The Creation of a Binaural Spatialization Tool

A PhD Thesis submitted in partial fulfilment of
the requirements for the degree of
DOCTOR OF PHILOSOPHY

Music, Technology and Innovation Research Centre
Faculty of Humanities
DE MONTFORT UNIVERSITY
Leicester, United Kingdom

by

Lorenzo PICINALI

November 2010

The Creation of a Binaural Spatialization Tool

A PhD Thesis submitted in partial fulfilment of
the requirements for the degree of
DOCTOR OF PHILOSOPHY

Music, Technology and Innovation Research Centre
Faculty of Humanities
DE MONTFORT UNIVERSITY
Leicester, United Kingdom

by

Lorenzo PICINALI

November 2010

Supervisors: Prof. Leigh Landy (MTIRC)
Dr. Dylan Menzies (Faculty of Technology)

Abstract

The main focus of the research presented within this thesis is, as the title suggests, binaural spatialization.

Binaural technology and, especially, the binaural recording technique are not particularly recent. Nevertheless, the interest in this technology has lately become substantial due to the increase in the calculation power of personal computers, which started to allow the complete and accurate real-time simulation of three-dimensional sound-fields over headphones.

The goals of this body of research have been determined in order to provide elements of novelty and of contribution to the state of the art in the field of binaural spatialization. A brief summary of these is found in the following list:

- The development and implementation of a binaural spatialization technique with Distance Simulation, based on the individual simulation of the distance cues and Binaural Reverb, in turn based on the weighted mix between the signals convolved with the different HRIR and BRIR sets;
- The development and implementation of a characterization process for modifying a BRIR set in order to simulate different environments with different characteristics in terms of frequency response and reverb time;
- The creation of a real-time and offline binaural spatialization application, implementing the techniques cited in the previous points, and including a set of multichannel(and Ambisonics)-to-binaural conversion tools.
- The performance of a perceptual evaluation stage to verify the effectiveness, realism, and quality of the techniques developed, and
- The application and use of the developed tools within both scientific and artistic "case studies".

In the following chapters, sections, and subsections, the research performed between January 2006 and March 2010 will be described, outlining the different stages before, during, and after the development of the software platform, analysing the results of the perceptual evaluations and drawing conclusions that could, in the future, be considered the starting point for new and innovative research projects.

Acknowledgements

I am particularly thankful to my supervision team, and to the whole MTI research centre, for the support given in the last four-and-a-half years. Thanks also to Dr. Brian Brown for the assistance given during the first subjective test, to Markus Noisternig, and to my uncle Andrea Ichino for their support during the second subjective test, and finally to Paddy Long for all that concerned my (English) written language.

Table of Contents

Preface

Before beginning with the “official” introduction to the thesis, some words should specify the target audience at which this work is aimed and therefore the level of technical knowledge required in order to understand what will follow.

The author, together with his supervision team, has decided to make this work available to a population larger than that comprised only of audio technology experts. For this reason, all of the technical terms, principles, formulae, issues, etc., of which knowledge is essential in order to appreciate all that is written between the introduction and the conclusion of the whole dissertation, have been carefully and simply explained in the first chapter, Chapter 1. Here, an introduction is given to the basic principles of acoustics, psychoacoustics, digital signal processing, and filter design in order to help the reader to grasp the concepts and theories that have been (in the introduction) and will be (in the rest of the chapters) outlined. It is therefore essential that a reader who is not particularly familiar with the basic knowledge mentioned in the previous lines starts reading from Chapter 1, passing then to the preliminaries, discussion, and the analysis offered in the thesis.

Introduction

The focus of this research work is, as the title suggests, binaural spatialization. The interest of the author in this particular field goes back to the final year project towards his undergraduate degree, when he first created a dummy head and began to implement simple software that would spatialize statically a sound source over headphones.

Binaural technology and, most of all, the binaural recording technique are not particularly recent (early examples of binaural recordings can be found, unexpectedly, in the last decades of the nineteenth century); it is only within the last ten to fifteen years that the increase in the calculation power of personal computers started allowing a complete real-time simulation of three-dimensional sound-fields over headphones. Here is exactly the reason why the author, captivated by the three-dimensional sound effect of simple binaural recordings realized with a hand-built dummy head microphone, decided to start a research project attempting to obtain with binaural synthesis the same spatial hearing effect that may be obtained with binaural recording.

It is worth now mentioning some of the progress during the four-and-a-half years that it took to achieve this final thesis: the Ph.D. work started in January 2006 at the Music, Technology and Innovation Research Centre, MTIRC, (Faculty of Humanities, De Montfort University, Leicester), of which the director is actually the first supervisor of the author. For the first nine months, the author worked full-time on the research project, before changing this to part-time after being appointed, in October 2006, to the position of Technician for the MTIRC laboratory. The author published in July 2006 his first paper, related to the topics of the Ph.D., at the DMRN 2006 conference in London (*see* Picinali, 2006). Furthermore, in April 2007 the author, together with his second supervisor, was invited to the Ear Club Seminar Series at the University of California at Berkeley to talk about the development of his research.

The research proceeded without any interruption until July 2008, when the author was appointed to the position of researcher in Paris, at the LIMSI-CNRS laboratory; he therefore left Leicester and started working in Paris on various research projects, always those linked with binaural simulation. He was simultaneously involved in a research project sponsored by GNReSound Italia[1] about the objective and subjective evaluation

1 *See* http://www.resounditalia.com

of the perceived quality of hearing aid devices. From June 2008, the author then began to work for IRCAM-CNRS, in Paris, on a research project related to data sonification and binaural spatialization.
In January 2009 he left Paris to return to Leicester, to the Faculty of Technology of the De Montfort University, where he is currently a Lecturer in Music/Audio Technology within the Department of Media and Technology.
In the following sections, this Ph.D. research work will be introduced and the structure of the thesis described in order to initiate the reader into binaural technology, and to allow him/her to enjoy the succeeding chapters with a wider knowledge of the facts described.
A further introduction to sound spatialization and the binaural technique is carried out in the last section of Chapter 1 (Section 1.7).

0.1 Goals of this body of research

It is essential, at the beginning of each piece of research, to review and analyse that which has already been done in a specific research field, and to establish the goals to be reached. This is exactly what is addressed in the following four sections.

0.1.1 About the state of the art

Within this Ph.D., extensive and continuing research has been carried out into the state of the art of sound spatialization, with particular attention paid to binaural spatialization. In the second chapter of this dissertation (Chapter 2, *The State of the Art in the Field of Sound Spatialization*), an overview of the outcomes of this particular stage of the Ph.D. work is provided. It is nevertheless important to underline in this introductory section that the outcomes of this state-of-the-art research provide a particular snapshot of what can be found on the market and of the world of research related to binaural spatialization. It was in fact found that none of the tested processors was able to deliver realistic and high quality binaural spatialization, and that a lack of precision and fidelity was often related to the following issues:

- Localization of the apparent image of the sound source inside the head, and therefore problems in the simulation of the distance
- Problems related to the localization of frontal sound sources
- Low quality of the HRIR database used

- Poorly implemented (or, more likely, totally absent) binaural reverb functions
- Absence of software for a direct conversion between the multichannel format and the binaural.

These outcomes (of which the origin will be described more precisely in Chapter 2) led to the development of the ideas underlying this whole research work, which can be summarised in one single sentence: to create an effective, efficient, and flexible binaural spatialization tool.

0.1.2 Convolution: Effectiveness and flexibility

An interesting parallel may be drawn between reverb processors and the binaural spatializator. When an algorithmic reverb simulator is compared with a convolution one, experts in the field of audio technology respond by agreeing with the fact that the latter are often more realistic, while the former offer much greater flexibility in terms of the acoustic characteristics of the environment to be simulated. This is perfectly explicable considering how the two types of simulation work: while the versatility of the convolution reverb is restricted by the actual impulse responses measured from the room to be simulated, it is possible with an algorithmic reverb virtually to simulate any kind of environment, even if this does not actually exist in the real world. Once the algorithmic model is set up, it can be fed with any parameters in terms of environment size and shape, absorbing the coefficients of the different surfaces, frequency response, etc. The convolution reverb would work perfectly if the room to be simulated perfectly corresponded to the measured one; however, if this statement is untenable, it becomes difficult to alter the impulse response in order to simulate environments with different characteristics. Indeed, convolution reverbs such as Altiverb[2] and Space Designer[3] allow the user slightly to modify the envelopes of the measured impulse responses, yet it is also the case that this is far from the flexibility provided by algorithmic reverbs.

A similar situation may be found when considering binaural spatialization. In this specific case, environmental simulations carried out using algorithmic reverbs, possibly performing the processing in the stereophonic or multichannel domain, converting then the signals to binaural using anechoic Head Related Impulse Responses, or HRIRs,

2 *See* http://www.audioease.com

3 *See* http://www.apple.com/logicstudio

cannot be considered particularly realistic. Further discussion of this appears in Chapters 2 and 8. However, they do offer maximal flexibility in terms of the characteristics of the environment to be simulated. On the other hand, binaural environmental simulations based on convolution with BRIRs (Binaural Room Impulse Responses) provide a superior level of realism, also as described in Chapters 2 and 8, yet have a very low level of flexibility; the level is even lower if compared with the standard (non-binaural) convolution reverb simulators referred to in the previous paragraph. It is not only a question of reverb simulation, but also of the preservation of the localization and distance cues, as are described in Chapter 5.

It is exactly in this scenario that the work presented in this thesis finds its place. The attempt is to create a binaural tool with an environmental acoustical simulation technique offering both the flexibility of an algorithmic reverb, and the realism of a convolution based one.

0.1.3 About head-tracking, individualized HRTFs and headphones calibration

As will be discussed in the following chapters, particularly in Chapters 3, 4, 5, and 6, various factors may influence the effectiveness of a binaural spatialization process. Three of these in particular are the implementation of head-tracking facilities; the use of individually measured or synthesized HRTFs (Head Related Transfer Functions), and the calibration of the playback system (mainly the headphones). Many studies into these topics have been made by different researchers (*see*, for example, Begault, 2001; McKeag, 1996; Bronkhorst, 1999, and Møller, 1996) and it may safely be stated that these factors will indeed enhance the realism of a binaural spatialization system.

However, implementation difficulties are predictable. The implementation of head-tracking facilities would require the use of sophisticated hardware tools in listening to binaurally spatialized signals, and could create problems related to the bringing to fruition of binaurally spatialized signals incompatible with the two-channel standard of the CD-DA, as well as to the complex calculations to be made before the actual playback. Added to these, the use of individualized HRTFs would also introduce significant complications regarding the measurement of HRIRs for each of the listeners. Implementing calibration routines for each pair of headphones possibly to be used by the listener would require additional processing to be implemented before the reproduction

of the spatialized signals.[4] These factors are neither the only ones, nor even the most important, as will be discussed in Chapters 3 and 5.

For this Ph.D. research, it was considered more important to develop an effective yet flexible and usable binaural spatialization tool than to implement indiscriminately all of the possible functions that could influence spatial hearing. It is commonly recognized that the best binaural spatialization effect obtained so far is achievable through simple binaural recordings[5], without head-tracking, individualized HRTF or playback system calibration, and therefore this is to be considered the level of realism to aim for when developing binaural applications.

0.1.4 Contributions of this research to the state of the art

This section will be much clearer to the reader after having read the whole thesis;[6] however, it is important to describe briefly and immediately the contributions of this research to the state of the art in the binaural spatialization field:

- Development and implementation of a binaural spatialization technique with Distance Simulation, based on the individual simulation of the distance cues and Binaural Reverb, based on the weighted mix between the signals convolved with the different HRIRs and BRIRs sets (*see* Chapters 5 and 6).
- Development and implementation of a characterization process for modifying a BRIR set in order to simulate different environments with different characteristics in terms of frequency response and reverb time (*see* Chapter 6).
- Creation of a real-time and offline binaural spatialization application, implementing the techniques cited in the previous points, and including a set of multichannel- (and Ambisonics) to-binaural conversion tools (*see* Chapter 7).
- Performing of a perceptual evaluation stage to verify the effectiveness, realism, and quality of the techniques developed.

4 Considering the calibration of the listening system, it also needs to be borne in mind that the alterations brought by the use of one pair of headphones rather than of another are not direction dependent, and therefore do not depend on the position of the simulated sound source. This means that the spatial hearing system could rapidly adapt to these alterations, restoring the auditory spatial perception. Similar comments, yet with relevant differences, may be made for the use of non-individualized HRTFs. More about the plasticity of the spatial hearing system can be found in Parks (2004) and King (2001).

5 However, this system is far from being flexible, as will be seen in the following chapters.

6 References to the innovations of this research work will be made throughout.

- Application and use of the developed tools within both scientific and artistic "case studies" (*see* Chapter 9).

0.2 Chapter organization

In the following lines, a very brief summary is given of the topics addressed within the different chapters of the thesis. Table 1 offers an overview of the global organization of the chapters.

In Chapter 1 a general introduction is made in terms of digital signal processing and concepts of acoustics and psychoacoustics, focusing on the specific topics related to the subjects involved in this research. An understanding of the topics outlined within this chapter is essential to the understanding of the contents of the succeeding chapters.

Chapter 2 focuses on the extensive and continuing research that has been carried out into the state of the art of sound spatialization during the Ph.D., focusing on the consumer and professional markets, and on the research centres involved in topics related to binaural spatialization. Chapters 3 and 4 provide a summary of the mechanisms of spatial hearing for the perception of the angle (Chapter 3), on the techniques and performed experiments for the measurement of an HRTF, and on the simulation of the localization cues through convolution with an HRIR (Chapter 4). Similarly, Chapters 5 and 6 summarise the mechanisms of spatial hearing for the perception of the distance (Chapter 5), on the measurement and editing of BRIR sets, and on the development of the two most important techniques elaborated within this thesis: distance simulation and binaural reverb techniques (Chapter 6).

Chapter 7 describes and analyses the organization and implementation of the real-time and offline binaural processing software, while Chapter 8 focuses on the perceptual evaluations carried out during and after the research stages. Possible applications of the tool developed are described and analysed in Chapter 9, and a summary is given of various collaborations between the author and different musical composers, focusing on the benefits of the feedback gathered during this stage for the calibration and amelioration of the binaural algorithm and tool.

Finally, Chapter 10 provides the conclusions of the whole dissertation, summarizing the topics discussed in the previous chapters and proposing possible further (and future) additions to and developments within the research work.

Introductory chapters	Chapter 1: Basic Notions
	Chapter 2: The State of the Art in the Field of Sound Spatialization
Binaural phenomena and simulation for the perception of the angle	Chapter 3: Binaural Phenomena for the Perception of the Angle
	Chapter 4: Measurement of an HRIR Data-base
Binaural phenomena and simulation for the perception of the distance and of the acoustical environment	Chapter 5: Binaural Phenomena for the Perception of Distance
	Chapter 6: Distance Simulation and Binaural Reverb
Implementation, evaluation, and possible applications of the binaural tool	Chapter 7: The Binaural Spatialization Tool
	Chapter 8: Subjective Perceptual Tests
	Chapter 9: Possible Applications of the Tool Developed
Conclusions	Chapter 10: Conclusions

Table 1. Global organization of the chapters within the thesis.

Chapter 1

1. Basic Notions

In the following Chapter, an introduction will be given in terms of digital signal processing and concepts of acoustics and psychoacoustics, focusing on the specific topics related to the subjects involved in this research.

1.1 Glossary

Before beginning the overview of binaural phenomena, of the psychophysiology of the spatial hearing system and of the simulation of three-dimensional sound fields, the definitions of a few terms need to be attempted, in order better to understand that which is to follow:

- Sound: it is the oscillation of pressure transmitted through a solid, liquid or gas; particularly, sound means those vibrations composed of frequencies capable of being detected by ears.[1] Sound is transmitted through gases, plasma and liquids as longitudinal waves, also called compression waves, which are waves of alternating pressure deviations from the equilibrium pressure.
- Sound event: the physical aspect of the phenomenon of hearing, such as a sound wave or a sound source.
- Sound localization: the judgement on the specific location of a sound source.
- Sound lateralization: it is feasible, most of all while listening to sound through a pair of headphones, that a listener is unable to localize sound sources outside the head, but to perceive the sound as coming from inside the head, with sound sources placed along an imaginary line starting at one ear and crossing to the other. This phenomenon is known as *sound lateralization*.
- Localization cues: specific attributes of the sound event that are used by the hearing system in order to establish the position of a sound source in a 3D soundscape.
- Monaural: relating to or involving a sound stimulus presented to one ear only.
- Binaural: relating to or involving a sound stimulus presented to both ears simultaneously. The word ‘binaural’ is often used referring to a 3D audio spatialization technique using standard stereo headphones as the reproduction system.

1 *The American Heritage Dictionary of the English Language*, Fourth Edition (2006). Houghton Mifflin Company.

- Transaural: it is often used to refer to different audio processing techniques allowing the reproduction of a binaural stereophonic audio stream (created using binaural spatialization processing) through a pair of loudspeakers (more information on transaural may be found in Section 2.6.6).
- Interaural: between one pair of ears.
- Diotic: relating to or involving a sound stimulus presented to both ears in exactly the same way.
- Dichotic: relating to or involving a sound stimulus presented to one ear differently from the sound stimulus presented to the other ear.
- HRIR: Head Related Impulse Response. It is the anechoic impulse response of the "head system" (head, pinna and torso), measured at the beginning of the ear canal for a given angle of azimuth and elevation, and for a given distance (more information about HRIR will be given in Chapter 4).
- BRIR: Binaural Room Impulse Response. It is the impulse response of the "head system" measured inside a room, or any other environment. It is basically the combination of a HRIR with a room impulse response.
- HRTF: Head Related Transfer Function, describing how a given sound wave input is filtered by the head, the pinna and the torso before the sound reaches the ear canal. It is often represented by an HRIR database, organized by azimuth, elevation and distance at which the HRIR have been measured. The terms HRTF and HRIR are often used as pseudo-synonyms, where HRTF stands for the transfer function of the "head system" represented in the frequency domain while HRIR stands for the same representation in the time domain.
- IHL: Inside-the-Head Locatedness. This describes the sensation, when listening to a standard stereo signal over headphones, of sound sources being lateralized inside the head of the listener instead of being properly localized outside the head.
- Contralateral ear: the ear that is situated on the opposite side from where the signal is coming.
- Ipsilateral ear: the ear that is situated on the same side as from where the signal is coming.

1.2 Spatial coordinates

In order correctly to localize a sound source in a 3D soundscape, a coordinate system needs to be established. Three planes need to be distinguished, each with the origin placed at the centre of the head (*see* Figure 1):

- Horizontal plane: placed at the superior margins of the two ear canals and at the inferior part of the ocular cavity.
- Frontal or vertical plane: placed at an angle of 90º to the horizontal plane, it intersects with this at the superior margins of the two ear canals.
- Median plane: placed at an angle of 90º to both the horizontal and the frontal planes, it constitutes the axis of symmetry of the head.

Using this system as a reference, the position of a sound source may be unequivocally defined by the *Azimuth* (φ, localization angle on the horizontal plane, calculated proceeding anti-clockwise), the *Elevation* (δ, localization angle on the median or frontal plane, calculated proceeding upwards), and the *Distance* (r, the distance between the sound source and the centre of the listener's head). In Figure 1, three sound sources are placed as an example; their spherical coordinates are:

- A: φ = 0º, δ = 0º, r = depending on the radius of the circles drawn
- B: φ = 345º, δ = 30º, r = depending on the radius of the circles drawn
- C: φ = 270º, δ = 0º, r = depending on the radius of the circles drawn.

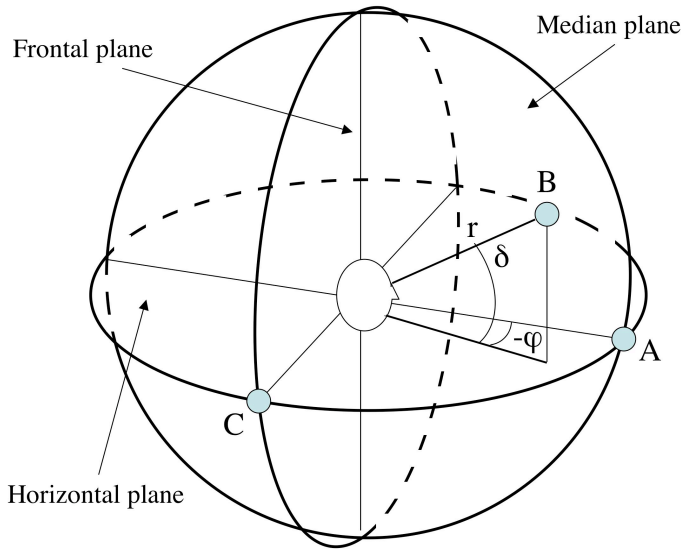

Figure 1. The spherical coordinate system (after Blauert, 1996)

1.3 Elements of the anatomy of the external ear

The human hearing system may be schematically divided into three parts: the outer ear (the pinna and the auditory canal), the middle ear (the eardrum and the three ossicles), and the inner ear (the vestibular and cochlear components). Regarding the mechanisms for the localization of a sound source, the outer ear certainly has a predominant function respect of the other two parts; for this reason, it will be analysed in detail in this section.

The outer ear is composed of the pinna (the visible part, *see* Figure 2), and the auditory canal or meatus.

After it has been conveyed and modified by the pinna (*see* Section 1.3.1), the sound travels down the ear canal and causes the eardrum, also known as the timpanic membrane, to vibrate. After this point, the vibrations are transmitted through the middle ear by the ossicles, three small bones (the *malleus* or hammer, *incus* or anvil, and *stapes* or stirrup) that work as impedance converters and mechanical amplification devices through a complicated system of levers, then to the cochlea, the last part of the auditory system and a component of the inner ear.

The hearing system as a whole does not, for our purposes, warrant scrutiny. The part of the peripheral auditory system involved in the mechanisms of sound modification linked to the source position, and therefore to the sound incidence angle, is in fact solely the external one, thus the outer ear.

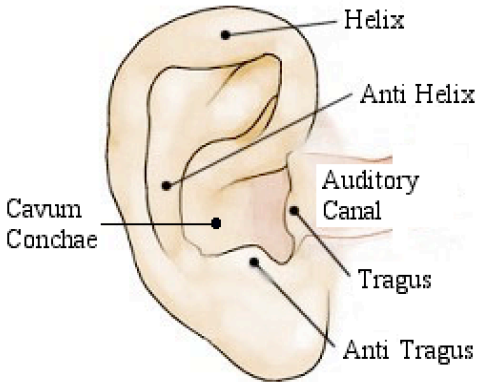

Figure 2. The external ear

1.3.1 The pinna

Composed of cartilage covered with skin, the pinna is located at an angle of 25°- 45° to the surface of the head; its physiological characteristics may substantially differ among

different individuals. At first sight, the role of the pinna seems to be quite simple: to convey the sounds reaching the head into the ear canal. However, if its particular shape is inspected carefully, its functions may be assumed to be far more complex. The pinna also in fact significantly modifies the incoming sound, depending upon the angle of incidence of the sound itself and thus on the position of the sound source. This modification is mainly related to frequency filtering, especially for high frequencies (usually above 3000 Hz), and it is known as 'Direction Dependent Filtering'; further discussion of this will be given in Chapters 3 and 5.

1.3.2 The auditory canal

The auditory canal is a slightly curved tube fully covered by skin. At the entrance it has a diametre of 5-7 mm, which then rises to 9-11 mm and diminishes again to 7-9 mm; its length is approximately 25 mm. It may be approximated as a constant diameter tube for frequencies up to 2 kHz (*see* Blauert, 1996, pp. 57-63), with high acoustic impedance walls. The propagation of the sound wave inside the canal may therefore be described by the mono-dimensional equation of a plane wave (given the fact that the canal has an average diametre of 8 mm, this approximation is valid for frequencies up to 23 kHz). Considering all these approximations, it is possible to offer the generalization that inside the canal there are no propagation losses.

The impedance of the auditory canal may be considered as equal in every single position in its insides: the calculation of its impedance is highly important, due to the fact that the auditory canal is strictly coupled with the pinna; this acts in this case as an impedance adaptor (exactly as does a Bessel tube for the trumpet), extending virtually the length of the meatus up to 30 mm.

During headphone listening, the propagation of the sound wave inside the canal happens in a substantially different way to that of a conventional listening situation, at least for some frequencies. It is generally believed that the sound pressure of the volume between the headphones and the eardrum is the same in every position. This may be considered as true only in respect of low frequencies; it has in fact been proven that for frequencies higher than 1000 Hz a real wave phenomenon appears within the canal (this is given by the relation between these frequencies and the length of the auditory canal itself; *see* Blauert, 1996, p. 54).

1.3.3 The eardrum[2]

The eardrum is an elliptical membrane (10-11 mm when measured at the long angle, and 8.5-9 mm on the shorter), approximately 0.1 mm thick, positioned at the end of the auditory canal at an angle of 40°-50°. It may be considered as a pressure sensitive receiver. It is caused to vibrate when there are pressure differences between the two sides of the membrane, one oriented towards the external ear and the other towards the middle ear. Of course, it must be assumed that the Eustachian tubes (linking the pharynx to the middle ear) are closed, and that the pressure inside the middle ear is constant.

The impedance of the timpanic membrane varies according to the various frequencies, and may increase up to 100 per cent, thanks to the 'Acoustic Reflex' phenomenon (*see* Blauert, 1996, p. 54), i.e., the contraction of two small muscles, located within the ossicles chain, activated when the sound pressure level reaches 90-100 dB.

1.4 Elements of DSP and filter design[3]

Digital Signal Processing (DSP) is the study of signals and of their processing methods in the digital domain; in contrast to Analogue Signal Processing, DSP is concerned with the representation of the signal by a sequence of numbers. In order to be more precise: a signal is any time-varying or spatial-varying quantity, and a digital signal is the numerical version of it, therefore it consists of a list of numbers, or a single number that changes with time or space.

The process allowing the representation of a signal by a sequence of numbers is called digitalisation. The numerical representation of a sound implies a loss of information from the perspectives of both frequency and the amplitude. The procedures for a correct signal digitalization are determined by the sampling theorem, which establishes the rules for a correct representation of the signal in the discrete temporal domain, and by the quantization theorem, that allows the representation of the sampled signal in the numerical domain with a finite precision.

2 Even if the eardrum is considered part of the middle ear, its location at the end of the auditory canal may justify a brief overview of its characteristics within this introduction to the physiology of the external hearing system.

3 For a far more complete overview on signal analysis, processing and filters design, *see* Rosenlicht, 1985; Frova, 1999; Rabiner, 1975; Openheim, 1975, and Cook, 1999.

Once a signal is correctly represented in the digital domain, it may be modified through a digital filter, a function that accepts as its input a set of one or more digital signals from which it generates as its output a second set of digital signals.

While an analogue filter works entirely in the analogue domain, and must rely on physical networks of electronic components (such as resistors, capacitors, or transistors) to achieve the desired filtering effect, a digital filter works by performing digital mathematical operations on a numerical signal.

Digital filters may be relatively complex, while those used in sound applications are often quite simple. Examples of digital (and analogue) filters may be produced referring to the low-pass and high-pass filters, the band-pass filter and the multi-band filter.

The question becomes more complex when analogue systems need to be simulated (*see* Farina, 2005, and Picinali, 2006).

1.4.1 Systems and transfer functions

Given two families of signals, F1 and F2, a system is an apparatus with the capacity to transform each F1 signal into an F2 signal. A system may be seen as a "black box", the behaviour of which is described by the transform law S: F1 → F2.

In environmental acoustics, a system is a room or a hall; in a recording studio, a system is an outboard effect; in an orchestra, a system is a musical instrument, all of which may be demonstrated schematically as follows:

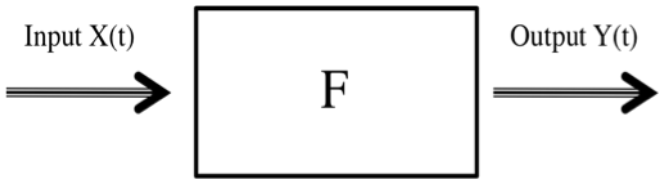

Figure 3. Schematization of a system

The mathematical expression to describe a system is:

$$Y(t) = F[X(t)] \quad [4]$$

It has now to be assumed that the system is:

4 This simply states that, being Y(t) the response of the system when input with X(t), Y(t) is in the function of X(t).

- Linear: the "overlap property" needs to be valid. If the input consists of a weighted sum of different signals, the output of the system is an overlap (that is, a weighted sum) of the replies of the system to all of the single signals in the input, as shown in the following formulas[5]:

$$X(t) + Z(t) \rightarrow Y(t)$$

$$Y(t) = F[X(t) + Z(t)] = F[X(t)] + F[Z(t)] \quad [6]$$

- Time-Invariant: it needs to be independent of time. If $X(t)$ is input and $Y(t)$ is output, for $X(t-t_0)$ the system has to output $Y(t-t_0)$, as shown in the following formula:

$$X(t) \rightarrow Y(t) \quad \Rightarrow \quad X(t-t_0) \rightarrow Y(t-t_0) \quad [7]$$

A particularly important notion about the linear spaces is its "basis". A basis is a set of arrays $\{a_1,...a_n\}$, and each array x may be obtained as a linear combination (weighted sum) of $\sum_i \alpha_i a_i$ elements of the basis, while at the same time no element of the basis may be obtained as a linear combination of the others.

A linear system S may be univocally defined knowing the responses of the system on the elements of a basis. In fact, for each input x, x is obtained as a linear combination $x = \sum_i \alpha_i a_i$, therefore:

$$S(x) = S(\sum_i \alpha_i a_i) = \sum_i \alpha_i S(a_i) \quad [8]$$

Knowing $S(ai)$, for each i, we are able to know $S(x)$ for all the x signals of the space.

---

5 Obviously, this is an approximate definition, yet it suffices in this case.

6 In order better to understand this equation, an example could be made through stating that a system performing a simple division may be considered as a linear system: if the system S1 performs a division by two, by inputting a 6, a 3 would result. 6 is equal to 4 + 2: the linear property says that inputting to the system 6, or inputting first 4 then 2, and summing the output, would generate the same resulting output.

In fact, 6 / 2 = 3, and (4 / 2) + (2 / 2) = 2 + 1 = 3.

7 Without delving into an explanation of the single terms of this equation, it would be sufficient to know that it simply proves whether a system is independent of time. An example could be made through stating that while a frequency equalizer may be considered a time-invariant system – because independently from the time when the signal is input, the output result will be the same – a dynamic compressor may not. In fact, inputting a signal just after a silence is different from inputting a signal after another signal with an amplitude peak at the end. In this second case, the possibility exists that the compressor is still in the 'release' mode, and therefore it would respond differently from if it were in 'standby' mode.

8 This equation is particularly similar to the second one: the response of a system inputting a sum of signals is equal to the sum of the responses of the system inputting the signals individually.

1.4.2 The impulse function, or Dirac δ

The function *δ(t)* may be thought as a rectangle with an "infinitesimal" base Δ and an infinite height 1/Δ (*see* Figure 4), so that:

$$\int_{-\infty}^{+\infty} p\Delta(x)dx = 1 \qquad [9]$$

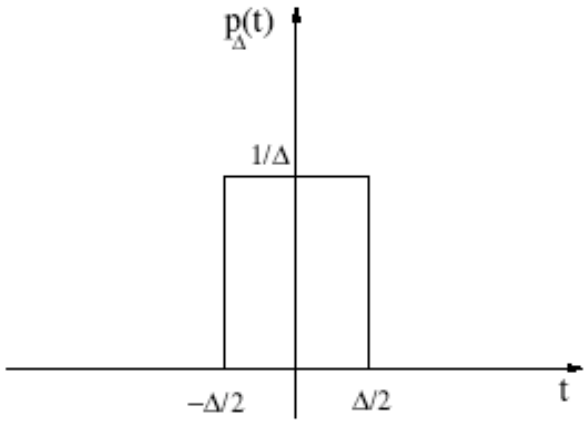

Figure 4. The Impulse Function, or Dirac δ

The frequency analysis of the Impulse Function is a horizontal line, parallel to the x axis. Therefore, it is possible to deduce that the *δ(t)* impulse contains all the frequencies at the same intensity.

It may be stated that a linear and time-invariant system may be described by its response to a specific signal: the Dirac Delta *δ(t)*. In fact, referring to the response of the system at the *δ(t)* impulse, it is possible to use as a characterizing element the function *h(t)*:

$$h(t) = F[\delta(t)] \qquad [10]$$

In a linear and time-invariant system, is it possible to describe the output *y(t)* with the following expression:

9 This equation simply states that the area subtended by the Dirac δ function is equal to 1.

10 This equation simply states that the transfer function of a system is equal to the response of the system at the Dirac δ signal. Therefore, inputting a Dirac δ into the system and recording the output would be the transfer function of that specific system.

$$y(t) = x(t) \otimes h(t) = \int_{-\infty}^{+\infty} h(\tau) \cdot x(t-\tau) d\tau \quad [11]$$

This formula is the so-called "analogue convolution".

1.4.3 The digital convolution

When referring to audio plugins for real time convolution (such as a convolution reverb), the reference domain is not analogue, but digital where, fortunately, the formulation of the convolution theory is particularly straightforward.

In the digital domain, the signals are represented dividing their variability interval into 2^n "sub-intervals" (this operation is called "quantization", where n is the number of bits used for the digital representation). The analogue signal is periodically measured (an operation called "sampling") and, depending on the value of the signal in that time gap, the sample takes on a value expressed in the number of n bits.

The signal enters the system as an array of numbers, and exits as another array of numbers, with the same sample rate and the same bitrate.

It is important to underline that the numbers in the output are directly related to the numbers in the input. For example, having input a sequence of zeros (silence) followed by non-null numbers and by zeros again, the output would result in a sequence similar to that in the input, although with a different number of zeros before and after the signal. This imparity is due to the fact that the response of the system is not immediate either when the system is excited (attack) or when the system goes back to its beginning state (release). In mathematical terms, it may be said that X_n is not just in the function of Y_n, but of a certain number of samples in input, starting from the n one and going backwards.

In the digital domain, this is expressed by the following equation:

$$y_n = (x_n * h_1) + (x_{n-1} * h_2) + (x_{n-2} * h_3) + \ldots + (x_{n-m} * h_m) \quad [12]$$

11 Simply stated, it is sufficient to know that this is the analogue version of the convolution operation, which expresses the amount of overlap of one function x over another function h (see http://mathworld.wolfram.com). The convolution is defined as the integral of the product of two functions after one is reversed and shifted. In simple terms, the signal x(t) is decomposed into simple additive components, and the response of the system to the input signal is obtained by adding the output of these components passed through the system itself.

12 This is the digital convolution; its operation is clearly shown in Figure 5.

where *m* is the last sample in the memory.

This operation is the digital convolution, and is indicated by the following expression:

$$y = x \otimes h \quad [13]$$

Therefore, the *h* coefficients are "characteristic" of the system. Viewing them as a waveform, they represent the impulse response of the system.

In order to provide a clearer and simpler example, Figure 5 represents the digital convolution process between an input array of 4 samples (X_1, X_2, X_3 and X_4) and an impulse response of three samples (H_1, H_2 and H_3); the output array would then be of (4 + 3) – 1 = 7 samples (Y_1, Y_2, Y_3, Y_4, Y_5, Y_6 and Y_7).

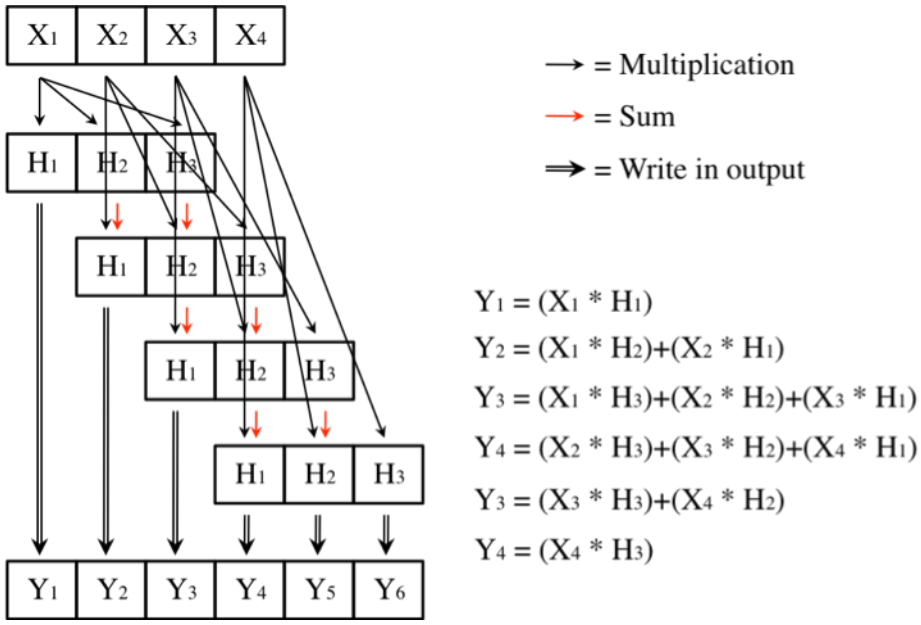

Figure 5. A schematic representation of the digital convolution process

1.4.4 IIR and FIR digital filters

Before proceeding to the subsequent chapters, a final distinction needs to be drawn between different types of digital filters. Depending on the nature of their impulse response, digital filters may be classified into two fundamental typologies (*see* also Figure 6):

13 The ⊗ symbol is the mathematical expression of the digital convolution.

- FIR (Finite Impulse Response) Filters. These have no direct correspondence with the analogue model and therefore they may be realized only in the digital domain. Their impulse response signal has a finite duration.
- IIR (Infinite Impulse Response) Filters: these are directly derived from the analogue model, and they are characterized by an infinite duration of the impulse response signal. They comprise a FIR filter plus a retroaction, or feedback, typical of all analogue filtering systems.

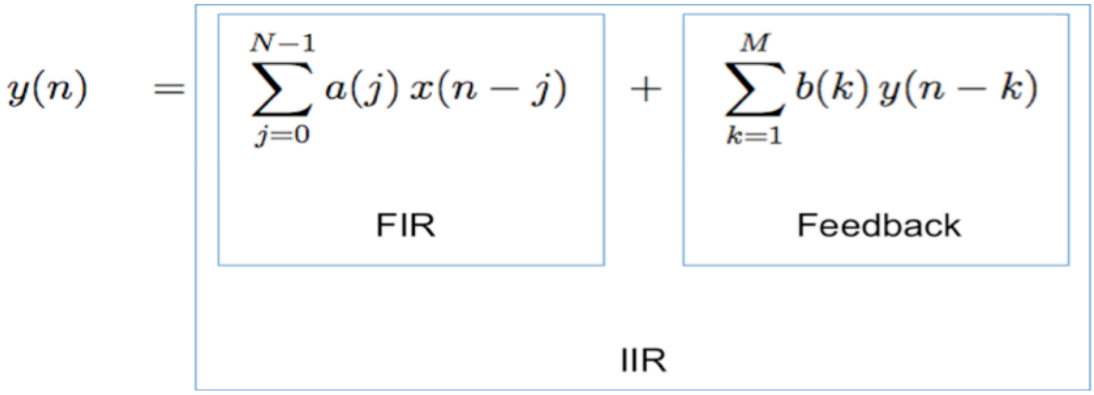

Figure 6. The typical model of FIR and IIR filters [14]

Nevertheless, with a certain number of approximations, it is possible to simulate an analogue filtering system with a Finite Impulse Response filter. The convolution itself is an FIR filter.

1.4.5 The Fourier analysis

The analysis of a sound allows to get into the informative microstructure of the signal and to obtain its analytical representation. The informative components of the signal,

14 It is not essential to understand all of the mathematical operators and functions of these equations. The sense of this representation is to explain that an IIR filter is composed of a FIR filter plus a feedback line. In order to make an example, a 'tape delay' filter may be considered (a filter which simulates, for example, the repeated echo effect reported when producing a loud scream or a whistle on the top of a mountain): this type of delay is composed of a simple gain reduction plus a feedback, that is, the reduced signal coming back to the gain reduction section after a predetermined amount of time (delay). If a signal is input into the system, it would then be repeated with decreasing amplitude at a rate given by the time delay value. In this case, the gain reduction may be considered as a FIR filter, with the whole 'tape delay' as an IIR. Continuing recursively to dividing the input signal by a given number, it would be impossible to arrive at the 0 value, thus the Infinite Impulse Response system.

measured through different analysis techniques, provide the basis on which extrapolating models for the modification and the synthesis of the acoustic information may occur.

The Fourier analysis is probably the most important of the frequency analysis techniques for audio signals, both because of its closeness to the sound perceptive model and of its relative simplicity in terms of mathematical model, a simplicity leading to a straightforward, fast application in the digital domain. The dynamic variability of both musical and vocal audio signals enforces the adaptation of the stationary conditions for the validity of the Fourier analysis to the dynamism of the audio signal: the Fast Fourier Transform[15] is an example of a dynamic adaptation of a stationary analysis model.

Following the Fourier harmonics analysis theory, complex signals may be broken down into a series of elementary sinusoidal signals, each with a different amplitude, frequency and phase. This de-composition is unique, and may therefore be used to code the information of the signal into another domain, different from the temporal: the frequency domain.

The pure tone (sinusoid) is the simplest example of audio information because it is characterized by a single frequency; every other signal without the informational characteristics of the pure tone is a complex tone and, according to the Fourier analysis, may always be broken down into different pure tones, with different amplitudes, frequencies and phases. If the complex tone is then a periodic oscillation, its frequency components exist only at the correspondence with the multiples of the fundamental frequency determined by the repetition period of the complex tone: the first harmonic (called also 'fundamental') is defined as the first pure tone with a period equal to the repetition period of the complex tone, the second harmonic has double frequency in respect of the first, the third harmonic has triple frequency, etc.

The starting point for the Fourier analysis is the Fourier series. This allows the calculation of the series of amplitude coefficients of the harmonic components of a continuous and periodic signal. Periodic signals are characterized by a waveform that is repeated, always identically, for the whole duration of the signal itself. In other words, the sinusoid is a periodic signal.

15 Further explanation follows.

'Real' signals are never perfectly periodic (periodicity is a mathematical abstraction), and for this reason the Fourier series is not applicable to them.

Nevertheless, the Fourier series has an equivalent, known as the Fourier transform, applicable to a-periodic signals. The Fourier transform is an extension of the Fourier series that considers the oscillation period of a signal as having an infinite duration.

The Fourier transform allows the calculation of the amplitudes and the phases of the harmonic components of a signal, which does not need to be periodic, for all frequencies, from 0 to infinity.

Figure 7 shows the mathematical representations of both the Fourier series and the Fourier transform.

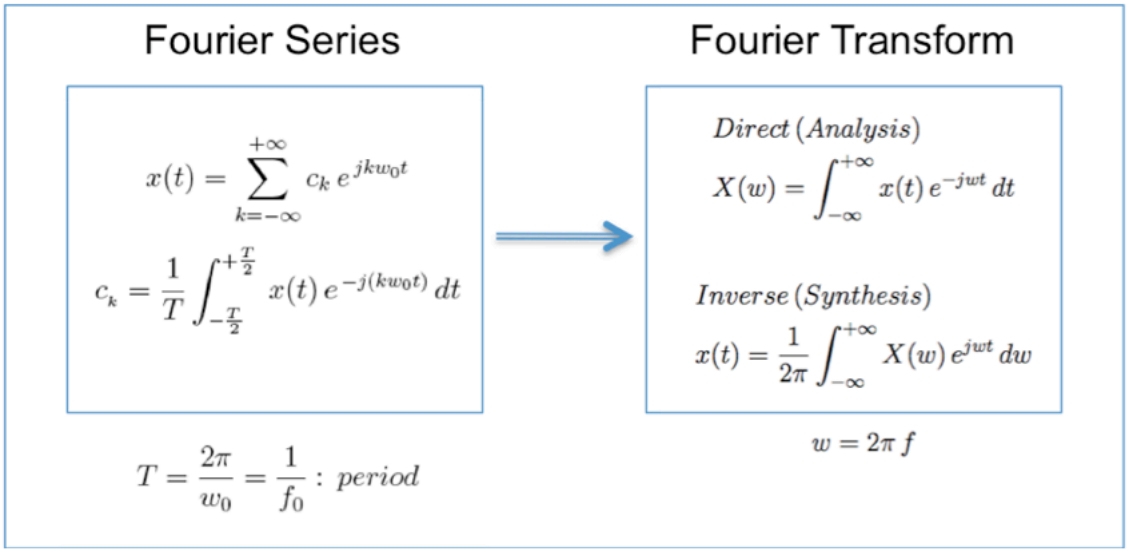

Figure 7. The Fourier series and transform [16]

As a simple reminder, the following line shows the Euler formula, which allows the passage from the Cartesian complex plane to the polar:

$$e^{jwt} = cos(wt) + j\ sin(wt)$$

16 Even in this case a precise analysis of the mathematical operators and functions will not be attempted. Nevertheless, it is important to understand the general meanings of these equations. In the first section, the 'Fourier Series' one, it is stated that each periodic signal is composed of a sum of pure signals (sinusoids) with a given amplitude and phase. In the second section, the 'Fourier Transform', two different subsections are outlined: the 'Direct (Analysis)' explains how to pass from the time domain (the periodic complex signal changing in time) to the frequency domain (frequency, amplitude and phase of all of the pure signals composing the original complex one), while the 'Inverse (Synthesis)' explains how to return from the frequency domain to the time domain, doing exactly the opposite of the operation explained before. The first operation is known as the Direct Fourier Transform (Analysis), and the second as the Indirect Fourier Transform (Synthesis).

As has already been outlined in Section 1.4.3, in working with computers the reference domain is digital. The Fourier series and transform described previously refer to continuous signals, while in the digital domain there are only discrete (sampled) ones. For this reason, a particular operation may be performed in the digital domain: the Discrete Fourier Transform (also known as 'DFT'), which works exactly as the Fourier transform, but with sampled signals.

1.4.6 The Fast Fourier Transform, or FFT

A more efficient implementation of the DFT has been created in order to speed the calculation time of the algorithm in the digital domain. This is called the Fast Fourier Transform, also known as FFT.

The fundamental concept forming the basis for the calculation speed of the FFT is that a DFT may be deconstructed into a greater number DFTs when applied to increasingly smaller signal portions. The calculation time requested by a DFT performed on the whole signal is greater than that requested by various DFTs performed on smaller portions of the same signal. In order to make the FFT algorithm even more efficient, the size of the time sections of the signals to which the DFT is applied, measured in samples, needs always to be a power of two. In fact, through exploiting intrinsic properties such as symmetry, the calculation model may become much faster.

The perfect periodicity of a waveform is an abstract concept. It is not possible to produce a sound perfectly stable in terms of frequency and amplitude. The sinusoid is therefore only a theoretical model. Nevertheless, a waveform with evident elements of periodicity may be considered as periodic even if there exist minor variations in terms of the oscillation period or waveform amplitude.

In order to analyse a real signal, it is essential to have a temporal segment of the signal itself, of which the length will be significant for the precision of the calculations. Sound analysis is based on the hypothesis of a stationary spectrum, a hypothesis implying a signal with a perfectly periodic waveform, with a period conforming to the length of the signal segment considered for the analysis itself. As this cannot be true because, as has been stated, the perfect periodicity of a signal is merely an abstract concept, approximations need to be made, and the analysed signal could result in being more or less different from the original signal.

A signal on which an FFT is performed needs to be divided into temporal sections, named “windows”: the dimension of the windows is very important for the frequency and time resolution of the analysis. Using smaller windows would increase the time resolution yet decrease the frequency response range, and *vice versa*.

This windowing process implies artefacts that may produce differences between the results of the analysis and the real characteristics of the signal itself; the distortions of the frequency information come as a direct consequence of the abrupt cut operated on the signal in order to obtain the time windows. The implicit stationary hypothesis leads to a distortion of the waveform, which then leads to a parallel distortion of the measured spectrum.

Figure 8 shows a schematization of the windowing process and of the distortions brought by the approximation of the periodicity of the windowed signal.

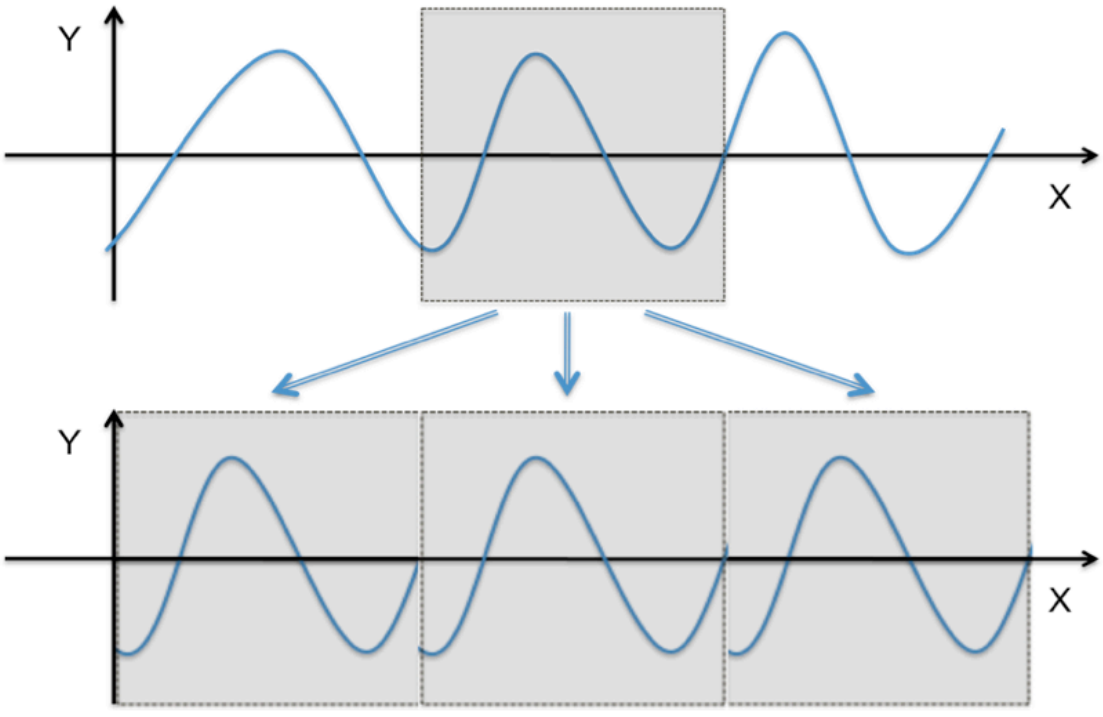

Figure 8. Schematization of the windowing process and of the distortions brought by the stationary hypothesis approximation

The windowing operation consists of a product between the signal to be analysed and a particular rectangular waveform signal with minimum null (0) and maximum unitary (1) amplitudes. This signal (window) is a unitary impulse with finite duration, therefore with a sin x / x spectrum.

The product in the time domain corresponds to the convolution in the frequency domain; the spectrum of the window is therefore propagated on each frequency of the windowed signal, producing a resultant spectrum consisting not of frequency impulses, as should be the case according to the Fourier harmonic analysis, but of a series of bells in correspondence with the position of the different frequencies within the spectrum. The windowing of a signal implicitly produces a series of artefacts on the real spectrum, the most relevant of which are the bell dilatation of the frequency impulses and the lateral oscillations, known as 'ripples', that appear on the sides of the bell (*see* Figure 9).

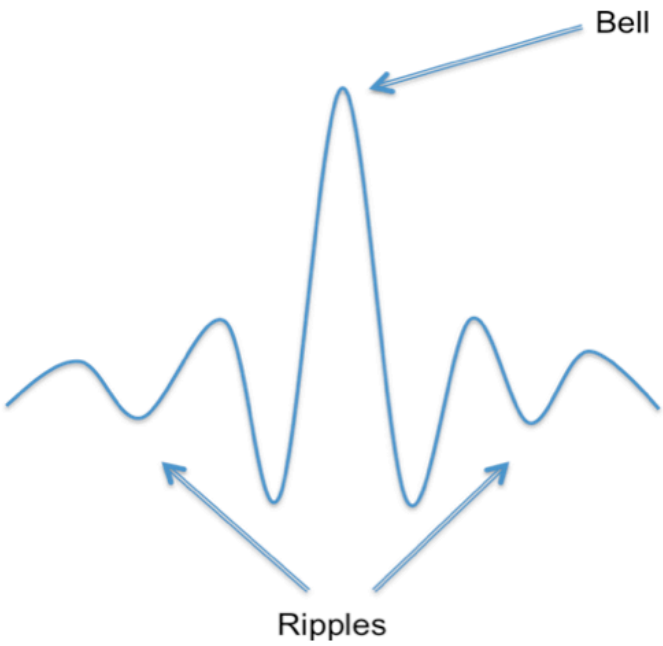

Figure 9. The bell in correspondence with the frequency impulses and the ripples

Different kinds of non-rectangular window functions may be used in order to minimize the ripples and shrink the bells created by the windowing process. These functions are known as 'cosine windows', characterized by decreasing amplitude at the extremes and

peak unitary amplitude at the centre. The increasingly lower amplitude at the sides minimizes the effect of the abrupt cut of the rectangular window. The Hanning, the Hamming and the Blackmann windowing functions may be cited as examples: [17]

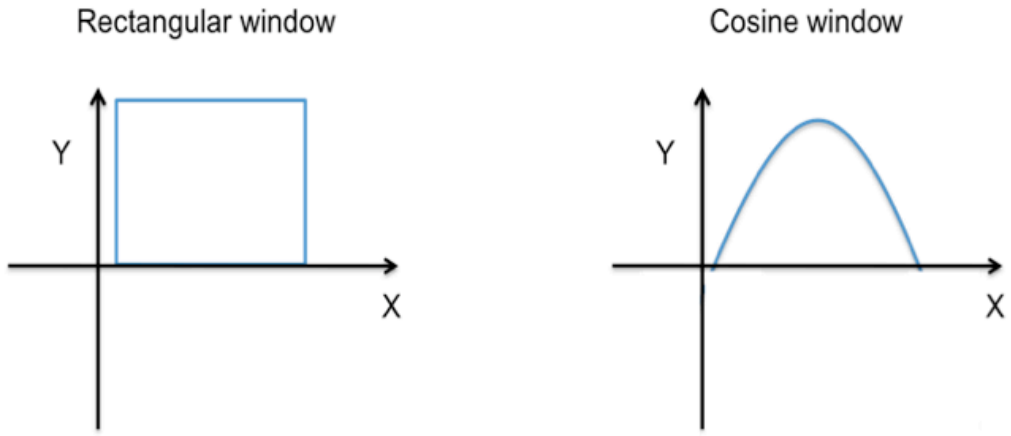

Figure 10. A graphical example of a rectangular and a cosine window

As in many other DSP applications where the signal needs to be treated in segments, within the FFT calculations so far described different factors need accurately to be considered. During the windowing process, the segments are affected by the convolution performed between the window function and the signal; thus, if an N sample segment is convolved with an M sample window signal, the output signal would be (N+M)-1 samples long. When recombining the different segments after the analysis, problems may occur in terms of time stretching if this lengthening of the convolved windows is not taken into consideration.

In order to overcome this problem, the 'overlap-add' method may be used. Given the example shown in Figure 11, a nine-sample input X(t) signal is decomposed into three segments X_i of three samples each. Each segment is convolved with a windowing function H(t), composed of three samples, and produces a segment Y_i (3+3)-1=5 samples long. When the three Y_i segments are then recomposed, a two-sample overlap needs to be considered in order correctly to obtain the Y(t) output signal:

17 *See* Rabiner, 1975, and Openheim, 1975

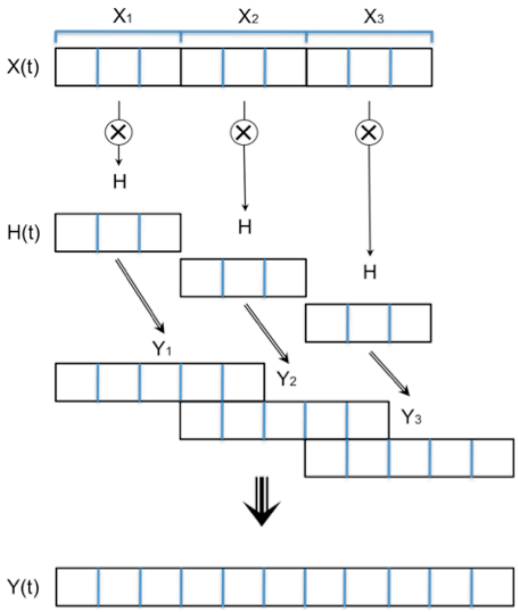

Figure 11. Schematic view of the overlap-add convolution method

Another method that could be used for the same purposes is 'overlap-save', which will not be addressed within this thesis.[18]

1.5 Different representations of an audio signal

Different signals may be represented graphically in different ways, depending on the domain (time and/or frequency) and on the information type that needs to be represented. The complexity of an audio signal may be significantly high, thus it is common practice to apply three different representations of the same signal:

- Oscillogram (time/amplitude): it plots the signal on a time/amplitude Cartesian diagram. With this representation, information may be obtained about the peaks and the amplitude of the sound wave, and an approximation of its periodicity and frequency content may be acquired.
- Spectrum (frequency/amplitude): it gives the amplitude of every frequency component for a given time-window. With this representation, information on the frequency composition of the sound wave in a specific moment may be obtained.

18 For more information on the overlap-add and overlap-save methods, *see* Rabiner, 1975, and Openheim, 1975.

- Spectrogram or Sonogram (time/frequency/amplitude): it gives the variations of the spectrum in time. Usually, the diagram is a simple two-dimensional Cartesian plane with time on the x-axis and the frequency on the y-axis, where the third parameter (amplitude) is represented by the colour (usually a gray-scale) of the section. Even a "waterfall" version of the spectrogram exists, representing the signal in a fully three-dimensional Cartesian diagram with frequency on the x-axis, amplitude on the y-axis, and time on the z-axis.

Figure 12 demonstrates a screenshot of a typical complete audio signal analysis performed with a well-known software for the audio analysis and modification.[19] In this specific case, the spectrum is represented in a vertical way (x-axis and y-axis inverted), in order to have the same y-axis of the adjacent spectrogram. Within the image, the three representations previously outlined may be distinguished. The signal is a flute A2 note; perceived are its periodicity within the oscillogram, its harmonics, with respective amplitudes in both the spectrum and the spectrogram, and the envelopes of each of its single components in the spectrogram.

Other representations of greater complexity and linked to a more complex analysis of a signal may be made, although will not be discussed here.

19 Audio Sculpt from IRCAM, (www.ircam.fr).

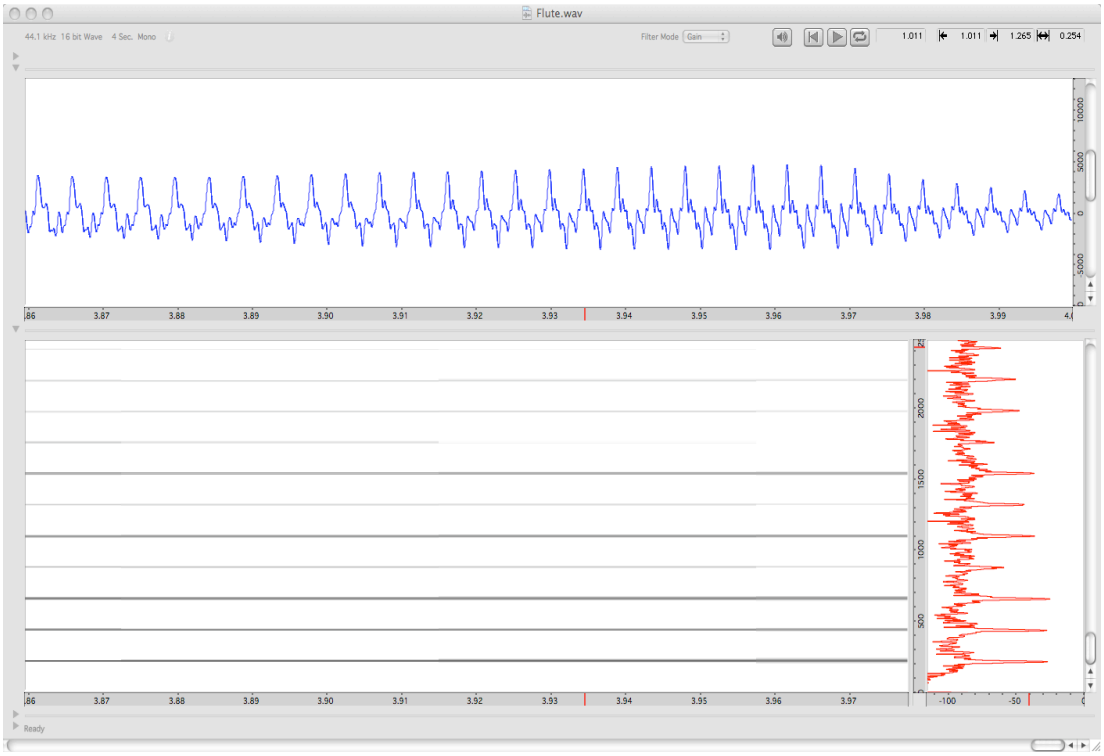

Figure 12. A typical complete audio signal analysis from Audio Sculpt (www.ircam.fr)

1.6 Elements of psychoacoustics[20]

The human ear is a complex system that transforms a sound event into an auditory event. Its behaviour is far from linear; no linear or proportional correspondence between the physical parameters of a sound and how it is perceived by the auditory system exists. As an example, physical parameters such as frequency and amplitude, absolutely independent upon each other in the acoustic domain, are linked within the acoustic perception domain.

Psychoacoustics represents the link between acoustics and cognitive psychology. It studies the relations between the acoustic phenomenon and the perception generated by it. Differently from acoustic principles, those of the psychoacoustic result from statistical data. Large numbers of individuals are asked about the perceived sensation when subjected to specific acoustic stimuli.

20 For a far more complete overview of introductions to psychoacoustics, *see* Frova, 1999; Moore, 2003, and Cook, 1999.

Various levels of audio signal elaboration contribute to the acoustic perception: for the outer and middle ear only in mechanical terms, for the inner ear in electrical, chemical and mechanical terms, and for the brain cortex only in electro-chemical terms. Furthermore, psychological components are often involved in these physiological transformations, making the psychoacoustic process even more complex.

Therefore, psychoacoustics cannot be considered as an "exact science", while acoustics is, yet it needs to be seen as a complex ensemble of physiological and psychological processes culminating in a specific perception of a sound event.

In order to provide an example of how psychoacoustics works as a link between the sound event and the auditory event, the four standard parameters of a signal may be cited then linked with the way they are perceived by the auditory system.

Generally, a signal may be analysed then described through the following four parameters:

- Amplitude: measured in sound pressure, volts, or some other units. Refers to acoustic signals, it is often expressed in terms of deciBels (dB).
- Frequency: measured in Hertz (Hz, cycles per second), it is the inverse of the period (T), which may be seen as the distance in time between two maxima.
- Duration: usually measured in seconds or milliseconds.
- Frequency content: the frequency, the amplitude (and envelope) and the phase (the position when the sine wave reaches its peak amplitude) of the pure tones that compose a specific signal (for further discussion of frequency content and Fourier analysis, *see* Section 1.4.5).

In Figure 13, amplitude and period are underlined within a sinusoidal wave time/amplitude representation (an oscillogram, as described in Section 1.5).

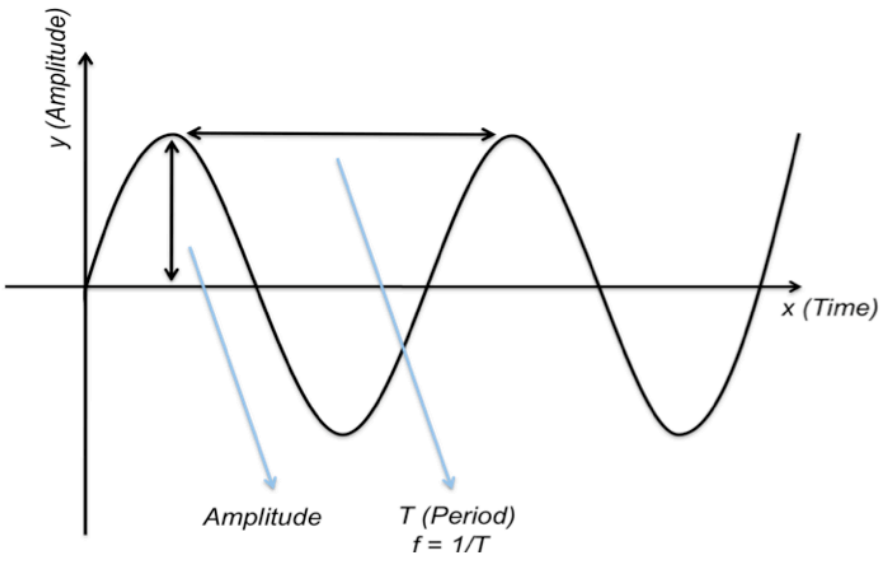

Figure 13. Time/amplitude representation of a sinusoidal wave

These four parameters belong to the physical (acoustic) existence of the sound (sound event), and they are, more or less, linked with parameters that belong to the psychological (psychoacoustic) perception (auditory event): Figure 14 shows how these links, which at a first sight could appear quite simple (see the blue continuous lines), become much more complex when analysed more carefully (see the red hatched lines).

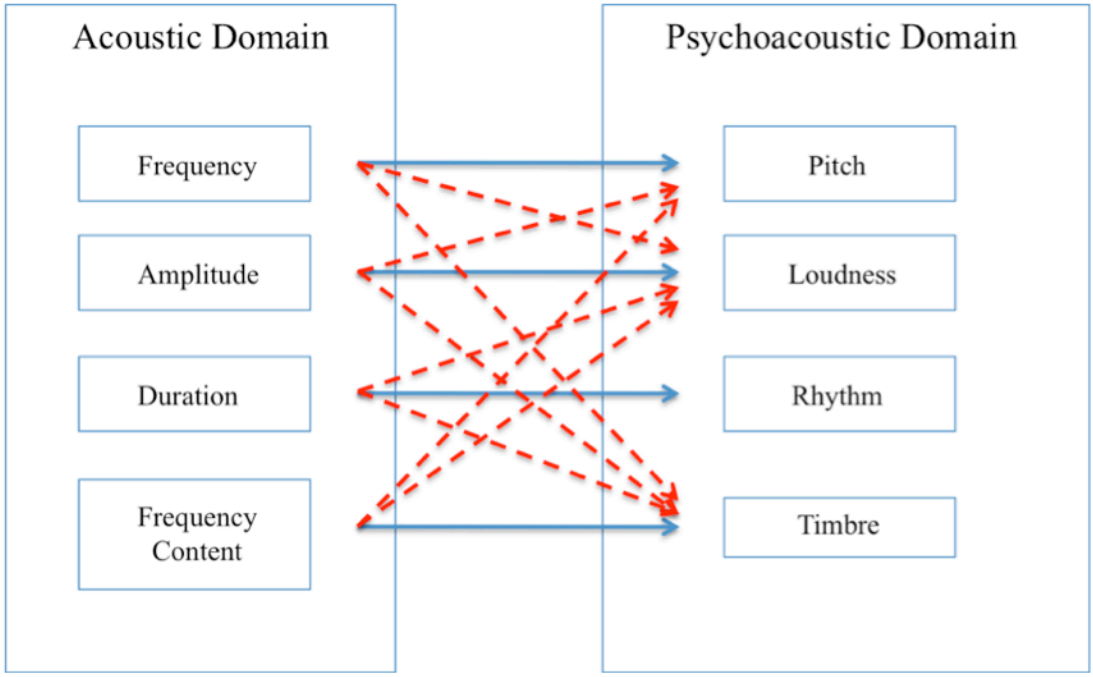

Figure 14. Schematization of the links between the acoustic and the psychoacoustic domains relative to the four standard parameters of a signal

Frequency is strictly linked with the perception of pitch: in the context of music, a direct connection exists between the frequency of a signal and a note. The same is true for the amplitude, which is connected with the perception of loudness. Duration is, in some ways, connected to rhythm, and the links between frequency content and timbre are far more complex than the others, and still strong.

Nevertheless, through performing a more careful analysis, it may be determined that the perception of pitch is influenced even by the amplitude and the frequency content of the signal, as well as the perception of loudness by frequency, duration and frequency content, etc. The links become far more complex and non-linear as the analysis deepens. A brief and simple overview of the perception of pitch and loudness will be performed in the following two sections.

1.6.1 The perception of pitch

Rather than going into depth into the physiology of the inner ear, it would be sufficient to comprehend that the structure of the cochlea, in particular of the cells on the basilar membrane, is organized in order to perceive the pitch. The non-linear mechanisms allowing the perception of the frequency of a sound are subjected, at a physiological

level, to a series of rules and limitations linked to the nature of the hearing system itself. As an example, the perceived distance between the 440 Hz and 880 Hz tones is the same as that between the 1000 Hz and 2000 Hz tones, and this is due to the particular organization of the cells on the basilar membrane.

Pitch is defined as the hearing sensation that allows the assigning of a sound to a determined position on a frequency scale (for example, the musical scale). The inferior limit of the pitch perception is the lower frequency at which the subject may still perceive a tone. This limit is subjective, although it is usually considered to be located between 16 Hz and 20 Hz. The superior limit is the highest frequency perceivable as a tone. Even this limit is subjective, and may substantially decrease with age and with hearing damage. A 20-year-old subject should be able to perceive tones up to 20 kHz, while a 40-year-old individual up to 15 kHz.

Another important aspect of the perception of pitch is its discrimination level: two tones at different frequencies are not always perceived as different. The ability to discriminate between two tones with different frequencies varies with the frequency of the tones themselves. The human hearing system is more capable in terms of pitch discrimination for frequencies between 500 Hz and 4000 Hz. This is due to the biological development of the hearing system linked to the frequency band of speech.

As has already been outlined above, even the duration, the amplitude and the frequency content of the tone condition the perception of pitch: all of these, added to various physiological limitations of the inner ear, make the phenomenon of the pitch perception highly complex, and thus it will not be investigated here.[21]

1.6.2 The perception of loudness

Loudness is the perceptive, and subjective, sensation produced by the amplitude of the sound signal. As applies to other psychoacoustic parameters, the link between the amplitude of the signal and the loudness perception is not linear.

The measuring unit for loudness is the phon: one phon is equal to the dB value produced by the sound pressure of a pure tone at the 1000 Hz frequency. Therefore, only

21 Further information about relevant literature on the perception of pitch may be found in Moore, 2003.

for 1000 Hz pure tones are the dB SPL[22] and the phon scales coincident; for the other pure and more complex tones, the situation becomes more complex.[23]

Figure 15 shows the Fletcher and Munson (*see* Fletcher, 1933), or Equal Loudness, curve (one of the most widely recognized graphics in audio engineering). On the x-axis is the frequency, and on the y-axis the amplitude (in this case, the Sound Pressure Level of the signal); the different curves correspond to the various phon values between 0 (the minimum audible sound) and 120 (the pain threshold).

This particular graph shows, at its most fundamental, that loudness perception is strictly dependent even on the frequency of the signal, and that the human hearing system is far more sensitive, in terms of loudness perception, to frequencies between 500 Hz and 4000 Hz. These data are, as is pitch perception, related to the biological development of the hearing system, and linked with the frequency band of speech.

Other scales exist for the measurement of loudness, yet they will not be analysed here.[24]

22 Sound Pressure Level (SPL) is a logarithmic measure of the sound pressure of a sound relative to a reference value (usually the minimum audible pressure variation). It is measured in decibels (dB), following the formula SPL = 20log10 (Prms/Pref) dB, where Prms is the sound pressure being measured, and Pref is the reference sound pressure (*see* also Frova, 1999, and Cook, 1999).

23 It is important to consider that 1000 Hz is not a particular coincidence between the sound event and the auditory event; it acts only as a reference point in order to be able to build a psychoacoustic measuring scale for loudness (*see* Figure 11).

24 More information about relevant bibliography on the perception of loudness may be found in Moore, 2003.

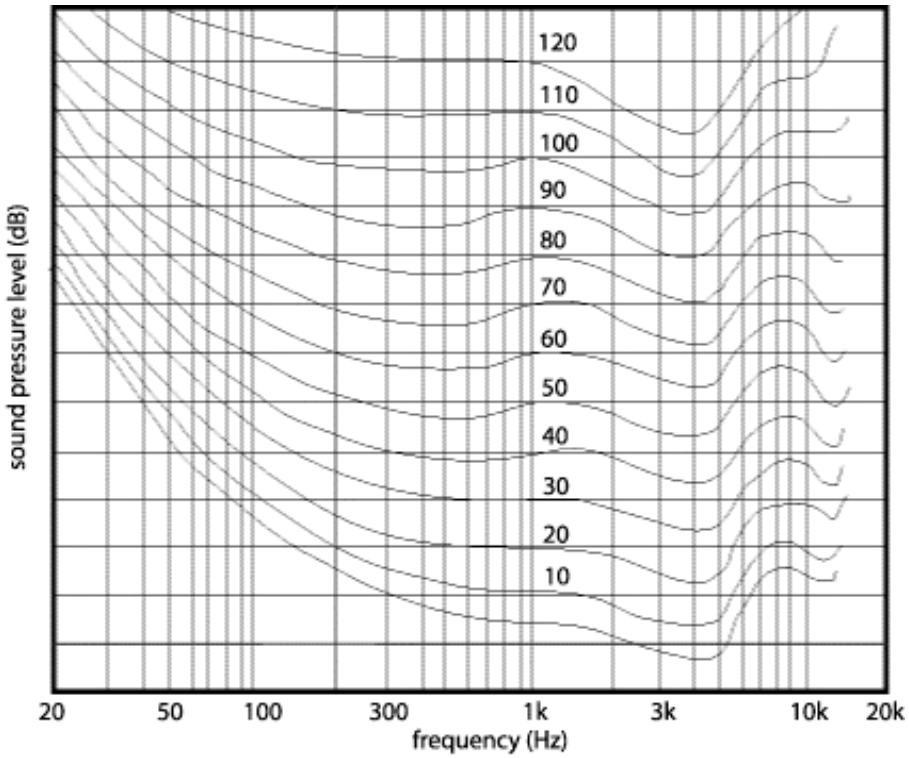

Figure 15. The Fletcher and Munson, or Equal Loudness, curve (after Fletcher, 1933)

1.6.3 Sound localization and space perception

In this thesis, links between other attributes of the signal, known as 'localization cues' (*see* Chapters 3 and 5), and the perception of the location of a sound source, or the perception of the spatial characteristics of the location where the signal has been reproduced, will be investigated; the aim of so doing is to synthesize those physical attributes and obtain the same perceptive effect as offered by real signals, when reproduced in a given position within a real environment.

1.7 Introduction to sound spatialization and the binaural technique

What does 'sound spatialization' mean? It could be considered as being related to synæsthesia, because the concept of space usually refers to the sense of sight, while the word 'sound' is, of course, related to hearing; nevertheless, these terms may be associated, and such an association creates a new concept: the soundscape. An attempt to define the concept of soundscape could start with a simple question: what is the difference between a listening with physical presence to the sounds that it is possible to hear

every day, for example, walking down the street, and the listening to a CD-DA (Compact Disc Digital Audio), played from any stereo reproduction system, of the same sounds recorded? Independently of the origin of the stimuli, in everyday life the sounds come from sources located in a 3D space: the listener is in the middle of an immersive 3D soundscape, where for each sound it is, more or less, possible to detect the position of its source. When listening, instead, to a CD-DA, the sound is presented frontally. Using a standard stereo reproduction set-up (with the two loudspeakers placed in two of the angles of an equilateral triangle and the listener placed in the third), the sound that reaches the hearing system is not 3D, but mono-dimensional, in the sense that each sound source can be localized in one or the other loudspeaker, or on a imaginary line between the two (someone might argue that a source can also be localized behind the loudspeakers, *see* also Section 1.7.1). In fact, a sound that is played at an equal level from both loudspeakers would be localized exactly between the two, thanks to psychoacoustic mechanisms that will not be discussed in this chapter (for more information, *see* Chapter 9, or Moore, 2003 and Blauert, 1996).

The main difference between the two listening situations has thus been defined: in the real one, a 3D soundscape is presented to the hearing system, while during a CD-DA playback the soundscape is mono-dimensional and frontal. In order to simplify the present discussion, the various interactions with the room where the CD-DA is played, interactions that can generate reflections coming from all directions and therefore stimulate the perception of a more spacious soundscape, will not be considered. Taking for a moment the playback of recorded sound: adding, for example, two loudspeakers behind the listener could help in coming closer to the experience of a 3D soundscape. If the four loudspeakers are placed at the corners of a square, with the listener located exactly in the centre, the sounds can be spatialized within a plane, thus a bi-dimensional soundscape can be created[25]. In fact, changing the weights (the levels) and the sound contents of the signals sent to the four channels, sound sources can be virtually located within the square described by the loudspeakers (again, for more information *see* Moore, 2003 and Blauert, 1996). This is called Quadraphonic reproduction system (Quad), and it was the starting point for more famous and recent surround systems such

25 Arguments can be brought saying that with a 90° angular separation between loudspeakers, it is rather difficult to virtually position a source anywhere within the loudspeaker's plane, but as outlined later in Section 1.7.1, these examples have been kept as simple as possible, to facilitate the basic understanding of sound spatialization techniques.

as Dolby Digital (5.1, 7.1, etc.) or THX Surround (*see* Chapter 2). Even if they more closely approach a proper 3D soundscape simulation, giving the impression of sound sources spatialized within a plane, they nevertheless lack one dimension, and they have very poor performances for side localization.

What if four other loudspeakers are added above, generating a cube with eight loudspeakers at the apexes and the listener placed exactly in the middle (*see* Figure 16)? Using this specific system a third dimension (height) can be simulated. Here follows some examples:

- If a sound is played at the same level from two frontal loudspeakers, the virtual sound source will be located in the middle of the line between the two loudspeakers. Through introducing differences in level between the two loudspeakers, the sound source can be moved into every position along that line.
- If a sound is played at the same level from four loudspeakers placed at the corners of a square, with the listener located in the centre, the virtual sound source will be located in the middle of the square described by the four loudspeakers (position of the listener). Through introducing differences in level among the four loudspeakers, the sound source can be moved into every position within that plane.[26]
- If a sound is played at the same level from eight loudspeakers placed at the apexes of a cube, with the listener located in the centre, the virtual sound source will be located in the middle of the cube described by the eight loudspeakers (again, the position of the listener). Through introducing differences in level among the eight loudspeakers, the sound source can be moved into every position within that space.[27]
- With multiple sounds played at different levels from the eight loudspeakers, a complex 3D soundscape can be generated.

26 Again, refer to Section 1.7.1

27 Again, refer to Section 1.7.1

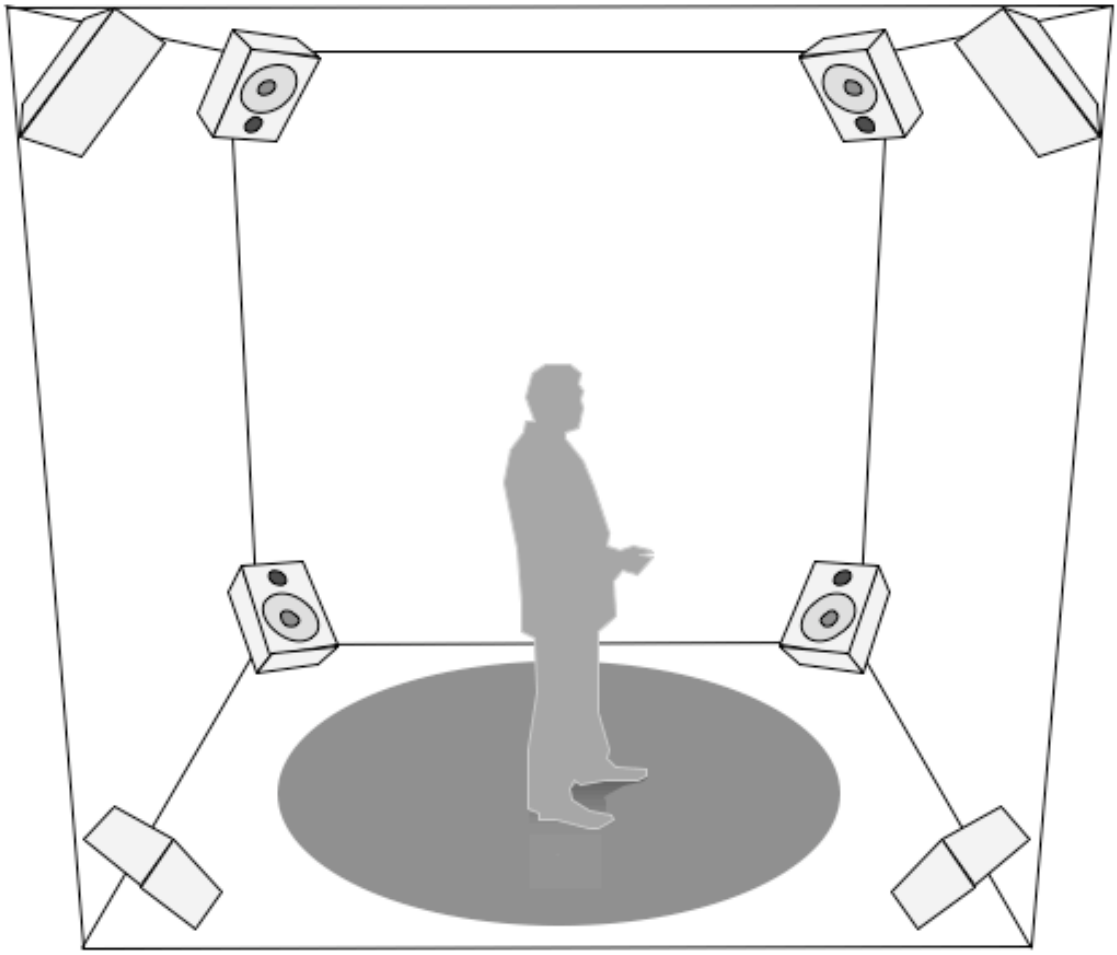

Figure 16. Eight-channel 3D audio reproduction system

Using this reproduction system with eight loudspeakers and, of course, a proper sound spatialization engine for the weighting of the respective signals in the eight channels, a 3D soundscape can be simulated (*see* Chapter 2). However, is this the only way to generate artificially a 3D soundscape? Eight loudspeakers, therefore eight channels, are used in order to be able virtually to locate a sound source in a 3D space, but – is not the hearing system made by only two ears, therefore two receivers?

In Chapter 3 and Chapter 5 of this thesis, the mechanisms of spatial hearing will be well described: sound sources can be located in three dimensions by using just two receivers. Thus, having sound reaching both ears by the use of a simple pair of headphones, it could be possible to eliminate complex and expensive multi-channel loudspeaker systems. Yet how can a three-dimensional soundscape be simulated when using only two channels?

The first and easiest way to simulate a 3D soundscape through headphones is simply to perform a binaural recording using a dummy head microphone or a pair of in-ear microphones. A dummy head microphone can easily be made by taking a head mannequin with the dimensions of an average adult human head, with sufficiently precise pinna reproductions (in order to preserve the information related to resonance, refraction and absorption typical of a human HRTF), and placing two miniature omni-directional microphones at the entrance of each of the auditory canals. The recordings made through placing this device in the middle of a 3D soundscape, then played back through a pair of headphones, will give the listener the impression of being exactly in the position of the dummy head, with sounds coming from every direction: left-right, front-back, and up-down. This result can be obtained even using the so-called 'in-ear microphones', which are simply two miniature omni-directional microphones placed inside the auditory canals of a subject, positioned at the entrance to the canal itself.

The fact that the microphones should be placed at the exact entrance of the auditory canal, and not at the position of the eardrum, may need some explanation: as happens with the pinna, even the auditory canal has its own resonances, while different studies (*see* Hammershoi, 1995) showed that these are not dependent on the angle of incidence of the input sound. Therefore, all of the localization cues are already present in the signal at the entrance to the ear canal, thus the microphone can be placed in that position.

A further observation needs to be made about the use of headphones for the reproduction of binaurally recorded sounds (and, as will be seen here, even for the reproduction of binaurally synthesized signals): when a stereo sound is played back through two frontal loudspeakers, the signal coming from the left loudspeaker will reach both the left and the right ears, exactly as will the signal coming from the right loudspeaker. This phenomenon is called crosstalk and, in the case of binaural sound reproduction, it would generate many unwanted situations. In fact, when playing binaural sounds, it is really important that the signal of the left channel should reach only the left ear, and that of the right channel only the right ear. Thus, the use of headphones is essential in this specific case: there exist systems that can be used to reproduce binaural sounds through stereo loudspeakers (transaural and crosstalk cancellation systems; as an example, *see* Tokuno, 1996), although they will not be discussed in this chapter.

The obvious problems linked to binaural recordings lie in the fact that the recorded 3D soundscape needs to be created in a real environment, using real sound sources or loudspeakers, and that the recorded scene cannot be modified after the recording. For these reasons, it cannot be considered a proper 3D sound simulation technique – simply a 3D sound recording technique.

In the following chapters, a proper 3D binaural tool will be presented, in order to be able to simulate 3D soundscape through headphones. In Chapters 3 and 5 the binaural mechanisms for the estimation of the direction and distance of a given source will be overviewed, and in Chapters 4 and 6 techniques for the binaural simulation of virtual sources in a 3D soundscape will be presented, moving finally the topic to the implementation of the actual binaural tool.

1.7.1 A short note

It should be noted that, in order to generate a bi-dimensional soundscape, it is not in fact essential to have four loudspeakers, as three placed at the corners of a triangle, with the listener in the centre, are sufficient; also, for 3D soundscape simulation, four loudspeakers placed at the apexes of a tetrahedron with the listener in the middle would be enough (even if the performances of such a system would be particularly poor). Furthermore, it might be argued that using a reverb simulation, the soundfield can be extended beyond the two, four and/or eight loudspeakers, and that with a very large angular separation between the different loudspeakers (~90°) it is difficult to obtain a high quality sound image.

All this is absolutely true, yet the attempt was to make the examples as simple as possible, and using an even number of loudspeakers seemed to aid clarity.

An overview of simpler and more complex sound spatialization techniques will be presented in Chapter 2.

Chapter 2

2. The State of the Art in the Field of Sound Spatialization

For this Ph.D., extensive and continuing research has been carried out into the state of the art of sound spatialization, with particular attention given to binaural spatialization. The importance of this stage can be summarised as follows:

- To justify the reason for the Ph.D. research, and to convince the reader of its originality;
- To establish the theoretical framework and the methodological focus of the research itself;
- To evaluate the approaches of each author-company-research group, and use each as a foundation on which to build the framework and planning of this research.

In this chapter, an overview of the outcomes of this stage of the Ph.D. will be provided. The first sections (Sections 2.1 and 2.2) will give information about the surround, virtual surround and binaural systems available for the consumer market (e.g., those implemented in DVD and CD players, standard software audio applications, and the consumer surround format). Section 2.3 offers an overview of multiple drivers surround headphones, while in Section 2.4 binaural and virtual surround systems (mainly software) for the professional market will be discussed. In Section 2.5, the results of a subjective quality evaluation test (performed by the author) on the most complete binaural systems available for both the professional and consumer markets are reported within a table, where more detailed information about the tested systems is offered.

In Section 2.6 different 3D sound simulation techniques (irrespective of their implementation) will be described; Section 2.7 (referring then to Appendix A) will recount specific researchers and research groups and projects working in the field of binaural spatialization. Finally, in Section 2.8 the research will be put into context, and the guidelines will be elaborated, taking into consideration that which has been accumulated from this preliminary research stage.

In the following pages, the reader may encounter technical words or expressions considered as commonly understood, thus not extensively explained within that specific section. Each of these provides a reference to Chapters 3, 4, 5 and 6, where explanations can be found. It is important to underline that many of the products and systems de-

scribed in these chapters are commercially available, and have been developed starting from registered patents; therefore, information about how the different algorithms and simulation work was either limited or even completely lacking, in certain cases: it has nevertheless been considered important to cite and list all these systems.

2.1 Surround formats in the consumer market

Surround sound refers to the use of multiple audio tracks and multiple loudspeakers to envelop the audiences watching a film or listening to music, causing the perception they are in the middle of a complex sound field that may, in the case of the movie or the music, represent the action or the concert. The surround sound formats rely on dedicated loudspeaker systems that literally and physically surround the audience. The position of the different speakers and the format of the audio tracks vary among the commercial companies specializing in this specific surround format. In the following lines a brief description will be offered regarding the most significant companies in the surround sound field, their formats, and their respective specifications.

Dolby[1]

> Founded in 1965, Dolby is one of the best known surround sound companies; its formats are probably the widest spread in the movie and audio industries, and it may justifiably be claimed that Dolby Surround set the standard in the surround audio field. (http://www.dolby.com)

An overview of their most famous surround audio formats is given, in chronological order, in the following list.

Dolby Surround

Dolby Surround was the "consumer" encoding of the Dolby multichannel analogue sound original formats for 35 mm films (Dolby Analog and Dolby Spectral Recording). From the mid-'Seventies, Dolby Labs began work on surround formats, especially for the cinema. The idea was to encode surround information on two channels, so that it would be possible to use the standard stereo recording media of those ages, along with, obviously, a surround decoder.

Dolby Surround is based on two channels, where four channels are encoded: front left, front right, centre, and rear surround; these were then decoded into Left, Right, Phantom Centre (resulting from the two frontal channels L and R), and Rear. The popularity of

1 *See* http://www.dolby.com

these kinds of encoding grew, and when they started to fall in price, they became a standard for the surround sound. The spatialization is performed according to the differentiation in the provision information about the audio content: the leading sounds come from the L and R channels, the speech from the Phantom Centre, and the environmental sounds and effects from the Rear channel.

This encoding works on matrix computation. Through the additions and removals of channels, the limiting of frequency bands, and phase modulations (it is similar to an M/S encoding performed for different frequency bands[2]) it encodes three channels into two (the fourth results from the L and R channels). The Rear surround channel is monophonic and has a limited frequency range (from 2 to 7 kHz). It can be said that Dolby Surround is a mixture of matrix encoding and compression reached through consideration of the different forms of information about the semantic content of the audio signal.

The real advantage of this format is that it can use the standard supports used for stereo files; the disadvantages are the single rear channel, which is also limited in frequency response, and the differences in frequency and phase among the three channels, nearly the same as in a stereo standard encoding.

Dolby Pro Logic

Dolby Pro Logic addresses the limitations of standard Dolby Surround by adding firmware and hardware elements in the decoding process. These have the capacity to emphasize important directional cues in a movie soundtrack; in other words, the decoding process will add emphasis to directional sounds by increasing the output of the directional sounds in their respective channels.

> This process, although not important in musical recordings, is very effective for film soundtracks and adds greater accuracy to effects such as explosions, planes flying overhead, etc. There is greater separation between channels. In addition, Dolby Pro Logic extracts a dedicated Center Channel that more accurately centers the dialogue (this necessitates a centre channel speaker for full effect) in a movie soundtrack. (http://www.dolby.com)

In addition to Dolby Surround, Dolby Pro Logic adds to the decoding chip hardware and firmware elements to emphasize the "sound directionality". The new chip can now

2 *See* http://en.wikipedia.org/wiki/Joint_stereo

add emphasis to the directional sounds (sounds differing among the three channels), providing them with greater loudness in the respective channel, resulting in a much more "surrounding" spatial perception. Even if this does represent a forward step from Dolby Surround, the Rear channel is still monophonic and has a limited frequency range.

Dolby Digital

As its name suggests, this system provides the digital encoding of a multichannel audio signal. It is not limited to a 5.1 format (Left, Centre and Right, Surround Left and Right, and the LFE or Low Frequency Enhancement channel); it can be extended to 7.1, 8.1, etc. The encoding of this format is done separately for each channel. In this way, the individually encoded channels share the same audio quality and the same frequency bandwidth (except for the LFE channel, which is limited to between 20 Hz and 200 Hz). The signal is then carried on a coaxial or optical cable (such as SPDIF or Tos Link, *see* Giesberts, 1995) through a multiplexing encoding (and demultiplexing on the decoder stage) for the digital organization of the signal. The actual encoding format for every single channel is AC3, a variation of the AAC MPEG format.[3]

Dolby Digital EX

It is a sub-class of the Dolby Digital format, developed in cooperation with Lucasfilm THX; the main difference from Dolby Digital consists in the adding of one more rear signal encoding two additional surround channels. This format is possibly the most complete within the Dolby family, and includes the following channels:

- Left Front
- Centre
- Right Front
- Surround Left
- Surround Right
- LFE
- Surround Back Centre
- Surround Back Right and Surround Back Left, encoded in the same channel.

It is fully compatible with Dolby 5.1.

3 *See* http://www.mpeg.org

Dolby Pro Logic II

Obviously, Dolby Digital and Dolby Digital EX encodings can be used only in conjunction with specific equipment, yet the compatibility is unilateral and is known as non-backward compatible or NBC. For this reason, Dolby Labs created a new Dolby Pro Logic type encoding, calling it Dolby Pro Logic II, based on matrix encoding, as is its predecessor; also, it can encode a 5.1 multichannel audio in a stereo channel (adding on channel plus a subwoofer channel to the classic Pro Logic).

Dolby Pro Logic II has also another function: it can extract a multichannel format from a standard stereo audio signal (with audio spreading techniques and phase modulations), although it needs to be stated that this function has never been positively evaluated by audiophiles.

There is also a Dolby Pro Logic IIx format, able to carry 6.1 and 7.1 formats in a single stereo channel.

Two other formats are present in the Dolby range, and they are Dolby Virtual Speaker and Dolby Headphones. Because of their nature (in the encoding and decoding process, major psychoacoustic simulations are performed linked with binaural spatialization), they will be described in Section 2.2.

DTS (Digital Theatre System)[4]

Dolby is not, of course, the only company that creates and commercializes surround sound formats; although mainly specialized in car audio systems, DTS Digital Entertainment has launched onto the market different formats that can be directly linked to the corresponding Dolby systems.

Basic DTS

It is a 5.1 encoding, similar to Dolby Digital, yet the actual signals are encoded using a minor compression ratio; this is done in order to guarantee a higher sound quality. Usually, although of course it is not a given, Dolby Digital is used for movies, and DTS for music.

DTS-ES

This is similar to Dolby Digital EX, but it has two different encoding typologies:

4 *See* http://www.dts.com

- DTS-ES Matrix: it creates the central rear channel with a matrix computation between the two rear channels of the 5.1
- DTS-ES 6.1 Discrete: the same as Dolby Digital EX, with a central rear channel individually encoded.

DTS Neo 6

Similar to Dolby Pro Logic II and IIx, it encodes 5.1 and 6.1 audio into a standard stereo channel.

ITU Recommendations[5]

Although these are themselves not surround sound formats, simply international recommendations for the standardization of the positioning of loudspeakers in a surround sound setup, it has been considered worth citing the two more important examples within this overview.

- **ITU BS 775-1:** Multichannel stereophonic sound system with and without accompanying picture. Recommendation for speaker placement in a surround setup, based upon experimentation by the BBC.
- **ITU BR 1384-1:** Parameters for international exchange of multi-channel sound recordings with or without accompanying picture.

Lexicon Logic 7[6]

LOGIC 7 technology is a proprietary suite of surround algorithms developed and introduced by Lexicon (a company famous mainly for its reverb processors). LOGIC 7 technology is used for recording and distributing multi-channel sound on two channel media. It uses psychoacoustic techniques to restore the original channel separation with very little change in sound.

The technology is based on a matrix encoding and decoding technique; it is designed as a two by n matrix, where n is the number of output channels. Each output can be seen as a linear combination of the two inputs, where the coefficients of the linear combination are given by the elements within the matrix.

A white paper explaining and analysing carefully this surround format is available on the Lexicon internet site.[7]

5 *See* http://www.itu.int

6 *See* http://www.lexicon.com/logic7/

7 *See* http://www.lexicon.com/logic7/whitepapers.asp

MPEG Multichannel and MPS[8]

Within the MPEG audio encoding range, it needs to be specified that MPEG-1, 2 and 2.5 encodings are merely audio formats, while MPEG-4, 7 and 21 are more complex encoding "frameworks" for audio signals, content information, video and other multimedia materials. The first ones implement lossy compression (Bosi, 2003) algorithms based on psychoacoustic principles and cochlear models, while MPEG-1 can encode a maximum of two channels, MPEG-2 and 2.5, and supports up to six channels (5.1).

Newer formats, such as AAC (Advanced Audio Coding), at the beginning called MPEG-2 NBC (Non-Backward Compatible), supports up to 48 channels, plus 15 channels for the LFE.

In the last three years, a new format called MPEG Surround (MPS) has been developed by the Fraunhofer IIS (the Fraunhofer-Institut für Integrierte Schaltungen,[9] the world's leading research group into compressed audio technology) and implemented in various consumer products and applications. In MPS a compact set of parameters representing the spatial image of the original surround signal, such as 5.1 or surround sound recordings, is transmitted, along with a mono or stereo downmix automatically generated in the MPS encoding process. The channel transporting spatial information is normally five to ten per cent of the dimensions of the downmix mono or stereo file, which is usually compressed using AAC, but can also support standard MPEG-2 compressions. It is important to underline that the MPS stream can be decoded with standard MPEG-2 or AAC stereo decoders, losing of course the surround spatial information while preserving the mono or stereo audio content. All of these factors make MPS a highly powerful and flexible surround compressed format.

Quadraphonic sound

Introduced into the American market in September 1970 as the Quad-8 or Quadraphonic 8-Track, "Quad" (as it became known) did not remain restricted to the discrete channel format used in the Quad-8. It appeared in several different and largely incompatible formats on different media, such as vinyl records, eight-track tapes, and reel-to-reel tapes. The Quadraphonic sound format was mainly divided into two sub-categories:

8 *See* http://www.mpeg.org

9 Fraunhofer-Institute for Integrated Circuits, *see* http://www.fraunhofer.de/EN/

- Four Channel Discrete: surround with four separated channels. It was prohibitively expensive for that age (the late 'Sixties), and there were no support regarding where to record, store and read four channels simultaneously.
- Quad: matrix encoding of a four-channel audio in a two-channel support. It was particularly useful, simply because it was possible to use the standard stereo recording supports, yet the amplifiers needed to have a Quadraphonic decoder. It is the predecessor of the Dolby Surround format.

SDDS (Sony Dynamic Digital Sound)[10]

SDDS is a digital film sound format comprised of the SDDS soundtrack, optically printed on both edges of 35mm film, and the SDDS playback hardware. It is designed exclusively for motion picture theatres, and there is no consumer equivalent. This format supports up to eight channels: Centre, Left Centre, Right Centre, Left, Right, Left Surround, Right Surround, and a full-frequency Subwoofer channel.

Spatializer Audio Laboratories Inc[11]

Spatializer Audio Laboratories Inc., a relatively small company when compared to Dolby or SRS, brought to the commercial market in 2003 the Spatializer enCompass AV™, a multi-channel audio enhancement technology that provides up to 6.1 channels of audio from a mono, stereo or matrix encoded source. Encoders and decoders for this specific format can be found in consumer and professional DVD players as well as in professional computer audio interfaces.

SRS Labs[12]

Even if SRS is better known for its Tru-Surround and HeadphoneTM transaural and binaural systems (which will be analysed in Section 2.2), two discrete surround formats are present within the product list of the company: SRS Circle Surround and Circle Surround II.

While the Dolby Digital and DTS approaches offer surround sound for a precise directional standpoint (specific sounds emanating from specific speakers), Circle Surround emphasizes sound immersion. To accomplish this, a normal 5.1 audio source is encoded down to two channels, then re-decoded back into 5.1 channels and redistributed back to

10 *See* http://www.sdds.com

11 *See* http://www.spatializer.com

12 *See* http://www.srslabs.com

the five speakers (plus subwoofer) in such a way as to create a more immersive sound without losing the directionality of the original 5.1 channel source material.

THX Surround EX[13]

THX Ltd is a company that certifies and gives support for the optimization of production and playback of entertainment content in the professional and consumer market. It is usually known for the certification of surround sound systems in cinemas.

THX also developed a patented technology called THX Surround EX, which has its roots in the Dolby Digital-Surround EX technology developed jointly by Lucasfilm THX and Dolby Laboratories. THX Surround EX is in essence a home version of the more complex THX-certified cinema surround systems, and it offers 6.1 channel decoding schemes based on bass enhancements, re-equalization of the signals and other techniques for the enhancement of the surround sound sensation. The new technology is backward compatible, with the ability to play any Digital 5.1 source in both the theatre and home versions.

2.2 Virtual surround, binaural and transaural techniques and systems in the consumer market

Continuing with a general overview of the state of the art within the spatialization field, in the following section attention will move to binaural and transaural techniques and systems. With its focus on the consumer products market, a description will be offered of the most famous and better known binaural and transaural systems available for home entertainment use.

This section is organized in a table with a schematic collection of information about the different systems. It is important to underline the fact that many of the systems listed in this table are based on simple stereo enhancement techniques (mainly originating in the the creation of phase and frequency shifts between the two, left and right, channels) which have nothing to do with real binaural spatialization algorithms. It has nevertheless been considered important to cite and list all of these systems. The table uses codes for the different system typologies, spatialization techniques, functions and environmental simulations in order to render it more compact and easier for rapid consultation. The table itself, with the legend, can be found in Appendix C.

13 *See* http://www.thx.com

2.3 Multiple driver headphones

Within this overview on surround and virtual surround sound technologies, it is surely worth mentioning the multiple driver systems for surround sound over headphones: these systems are based on the notion of incorporating multiple drivers positioned differently inside each of the two headphone's earcups, in order to give the impression of having sound sources placed in front or at the back of the listener's head. These systems are usually comprised of a box, which receives surround sound signals and, after matrix-based processing, returns a proprietary signal stream specific to the company that produces that system and usually composed of three signals for each channel to be sent to the headphones, and the headphones themselves. They implement no HRTF simulation (an exception can be found in Greff, 2008). Systems based on this technique started becoming popular in 2002, but after a few years, with the appearance of increasingly efficient technologies based on the binaural technique, they almost completely disappeared.

2.3.1 Firebox Medusa 5.1 Surround Headset (Speed Link)[14]

It is composed of a pair of headphones with four drivers for each and a controller station with the amplification stage, the DSP for the surround decoding, and various controls for parameters such as front/rear/centre volumes. It can be input with almost every surround format in 5.1.

2.3.2 Hear Force X-51, HPA and AXT[15]

This is a surround headphones system based on four drivers for each earcup (centre-front-rear-subwoofer). It can be input with a digital Dolby Digital 5.1 signal (coaxial cable).

2.3.3 LTB (Listen To Believe)[16]

LTB delivers a large variety of surround headphones, all based on the multiple driver system (three drivers for each earcup). The patented technology implemented in these products is called "Independent Speaker Chamber".

14 *See* http://www.templegames.co.uk/PC/Peripherals/Medusa-51-Surround-Headset-Speed-Link.asp

15 *See* http://www.turtlebeach.com/site/products/earforce

16 *See* http://www.ltbaudio.com/ltp-wr-51.html

2.3.4 Mentor Deluxe 5.1 (Sunnytech)[17]

This system is composed of a pair of headphones and a controller, with a USB connector for the direct interfacing with the PC. The headphone uses a patented technology called Six Audio Chamber Drivers, and it has six drivers for each earcup.

2.3.5 Zalman ZM-RS[18]

This system is composed of a pair of headphones (ZM-RS6F) and a controller for the surround decoding and the amplification stage (ZM-RSA). The headphones have three drivers for each earcup. It can be input with almost every surround format, and it outputs a six-channel patented surround format to be used with the ZM-RS6F headphones.

2.4 Virtual surround, binaural and transaural techniques and systems in the professional market

In Section 2.2, the focus of the general overview on binaural and transaural techniques and systems was oriented towards the consumer products market. The following section provides an overview focused on the professional market, thus on those advanced software and hardware systems that have been made available for professional sound engineers and researchers in the field of virtual surround and binaural spatialization.

2.4.1 AM3d Diesel Studio[19]

AM3d is a supplier of 3D audio technology for mobile phones and portable devices, car and home entertainment, and is “mission critical” (as defined in the internet site), or electronic self-protection and communication equipment. Diesel Studio is simply a demo program that allows the positioning of sound sources on a 3D space around the listener using a 3D positional audio technology based on HRTF simulation (no more information are available about this technology).

Diesel Studio uses also an interactive 3D audio sound engine called Diesel Power™, allowing on-the-fly positioning of sounds anywhere in the three-dimensional space

17 *See* http://www.pcextreme.net/mentor_deluxe.php

18 *See* http://www.zalman.co.kr/eng/product/code_list.asp?code=023

19 *See* http//www.am3d.com

surrounding a listener. Diesel Power™ can adapt to either headphones, or to two- or four-speaker systems.

2.4.2 Aristotel Digenis plugins[20]

Aristotel Digenis is an Experience Audio Programmer at CodeMasters[21] (an audio games production company) who has developed different surround and virtual surround libraries and plugins that are freely downloadable from his internet site. Here follows a brief description of the available plugins; references to the Ambisonic technique can be found in Section 2.5, and for the MIT Kemar HRTF library in Gardner, 1994.

- Ambisonic Bidules: it is a family of plugins for Ambisonic encoding, decoding, binaural decoding, and rotations, to be used within the Plogue Bidule software platform.[22] Within the Bidules plugins there is also the MIT HRTF Bidule, a processor for HRTF spatial positioning, therefore for the creation of virtual sound sources for binaural listening. The binaural spatialization is performed using a real-time fast convolution algorithm (Kiss FFT[23]) within the signal to spatialize it and the Kemar HRTF measured by the MIT.
- Amblib: a C++ library for Ambisonic encoding, decoding and rotations up to the Third Order
- MIT HRTF Library: an open source C-library making access to the MIT Kemar HRTF set through two simple functions.
- Charles Gregory's Bformat2Binaural: this Audiosuite plugin is simply hosted by Aristotel Digenis in his internet site, but it has actually be programmed by Charles Gregory. It is a plugin for the conversion between 2D First Order Ambisonic and binaural using an array of four virtual speakers placed on the horizontal plane.

2.4.3 Bauer (Stereophonic to Binaural DSP)[24]

It is a framework (bs2b, written on C and C++ and used for foobar2000, Winamp and Apollo audio player only) distributed under the GNU Lesser General Public License. It is compiled by Microsoft Visual Studio 2003, and tested on Windows XP systems. The

20 *See* http://www.digenis.co.uk

21 *See* http://www.codemasters.co.uk

22 *See* http://www.plogue.com

23 *See* http://sourceforge.net/projects/kissfft/

24 *See* http://bs2b.sourceforge.net

technology works on the simulation of the HRTF through IIR digital filtering: it converts stereo file to "binaural" file, using stereo spread algorithms, delays and frequency filtering with the goal of giving to the listener the sensation of a wider sound-field, with sources located outside the head. It works as a stand-alone executable programme, or as a plugin for Winamp, Winamp2, foobar2000 and Apollo digital audio players. The Bauer binaural technology is also supported by OpenAL.

2.4.4 CSound hrtfer Opcode[25]

It is a opcode (a function) for the binaural spatialization within the CSound programming environment.

These unit generators place a mono input signal into a virtual 3D space around the listener by convolving the input with the appropriate HRTF data specified by the opcode's azimuth and elevation values. hrtfer allows these values to be k-values, allowing for dynamic spatialization.

No information is given about the HRTF data used for the spatialization.

2.4.5 Edo Paulus (Eude) ep.binspat[26]

Edo Paulus is a Dutch sound artist who released a free binaural object for the MaxMSP visual programming environment.[27]

ep.binSpat~ is an msp object (abstraction) that handles binaural spatialization for use with headphones through the use of HRTF. It can virtually position a monaural source on the horizontal plane, 360 degrees around the listener's head, by dynamic control of two parameters: direction and distance. It does not control vertical positioning.

The HRTF data are the same as those used by Tom Erbe for SoundHack (*see* Appendix C).

2.4.6 Forum IRCAM Spat[28]

This is a complete 3D audio processor implemented as VST plugins, standalone application and MaxMSP object library, from the Spatialization Research Group of

25 *See* http://kevindumpscore.com/docs/csound-manual/hrtfer.html

26 *See* http://www.eude.nl/maxmsp/

27 *See* http://www.cycling74.com

28 *See* http://forumnet.ircam.fr

IRCAM[29] (Paris, France). Its central features are linked with the management of multiple loudspeaker arrays, yet it also has a section for the binaural spatialization performed with a convolution between the signal to be spatialized and HRIR extracted from the Listen HRTF database[30]. It also implements an advanced HRIR interpolation technique for moving sound sources, based on an independent processing of the ITD, in order not to create comb filtering effects when interpolating between HRIR of different positions on the horizontal and vertical planes.

A room simulation technique and a Doppler effect simulation are also implemented; it must be noted that the room simulation modules includes also parameters for the directivity of the simulated source. More about the IRCAM binaural technology will be given in Chapter 4.

2.4.7 Greg Schlaepfer Binaural Simulator[31]

Greg Schlaepfer is a musician and sound engineer. He recently released the Binaural Simulator, a VST plugin for the binaural virtual positioning of sound sources within a 3D space surrounding the listener, with room modelling reverb and crosstalk cancellation for transaural listening. The plugin is freely available in the internet;[32] however, no information is given regarding how the binaural spatialization and the room simulation are performed.

2.4.8 IEM Bin_Ambi[33]

IEM Bin_Ambi is a real-time rendering engine for virtual (binaural) sound reproduction, composed of a sophisticated object library for the Pure Data[34] visual programming environment. The majority of the library has been programmed by Markus Noisternig and Thomas Musil, within the IEM Sonevir sonification research project.

The theory that forms the basis of the proposed library is an improved Virtual Ambisonics Approach (Noisternig, 2003a and 2003b). Using this approach provides a computationally efficient implementation of virtual environments with:

29 *See* http://www.ircam.fr

30 *See* www.ircam.fr/equipes/salles/listen/

31 *See* http://www.gregjazz.com/index.php?page=resources

32 *See* http://www.gregjazz.com/download/BinauralSimulator.rar

33 *See* http://iem.at/Members/noisternig/bin_ambi

34 *See* http://www.puredata.org

- Multiple moving sound sources
- Room simulation
- Head tracking
- Time varying listener positions
- Interchangeable HRIR settings.

The proposed library provides a simple API (application programmers interface) to make it easy to use for scientific as well as for artistic projects.

Further discussion of the virtual Ambisonic binaural approach appears in Chapter 7.

Also from the same project comes AmbIEM, an implementation of an Ambisonics rendering system for SuperCollider 3,[35] largely developed by Thomas Musil and Christopher Frauenberger[36], and bin_ambi.OSC, a sound server application for real-time binaural audio rendering in Pure Data.

2.4.9 NASA-AMES SLAB[37]

SLAB is a real-time virtual acoustic environment rendering system originally developed in the Spatial Auditory Displays Lab at the NASA Ames Research Center. SLAB performs spatial 3D-sound processing allowing the arbitrary placement of sound sources in auditory space.

SLAB (Sound Lab) is a software-based, real-time virtual acoustic environment rendering system being developed as a tool for the study of spatial hearing. SLAB is designed to work in the personal computer environment to take advantage of the low-cost PC platform while providing a flexible, maintainable, and extensible architecture to enable the quick development of experiments. The software provides an API (Application Programming Interface) for specifying the acoustic scenario as well as an extensible architecture for exploring multiple rendering strategies. The SLAB Render API supports a number of parameters including sound source specification (waveform and signal generation), source gain, source location, source trajectory, listener position, listener HRTF (Head-Related Transfer Function) database, surface location, surface material type, render plug-in specification, scripting, and low-level signal processing parameters. For further information about SLAB, *see* Miller, 2001.

35 *See* http://supercollider.sourceforge.net

36 *See* http://sonenvir.at/downloads/sc3/ambiem/

37 *See* http://human-factors.arc.nasa.gov/SLAB/

2.4.10 OpenAL[38]

OpenAL is a cross-platform 3D audio API, mainly used with gaming applications. The library models a collection of audio sources moving in a 3D space that are heard by a single listener somewhere in that space.[39] OpenAL supports Bauer stereophonic-to-binaural DSP built-in for stereo output, and it is likely that in the future a greater number of binaural engines will be supported.

2.5 Quality evaluation tests

This section is organized in one extensive table where nine systems for the consumer and five for the professional market are carefully described and analysed; it includes information about a listening test carried out by the author, in the attempt to evaluate the effectiveness and reality of the spatialization performed by that specific software or device, reported schematically. The test has been carried out using two Genelec 1030 active loudspeakers and Fostex T50RP and Sony MD 7505 headphones.

As regards the consumer market products, nearly all of the demonstrations made were based on the comparison between standard stereo files and stereo binaurally spatialized ones. It needs to be said that the "standard stereo" files were often simple mono signals, with reduced frequency bandwidth (most or all of them with poor, low frequencies), while the spatialized signals were fully stereophonic, with clearly audible enhancement of the bass frequencies. Such comparisons could not then be considered utterly "fair", because of course a deeper and stereophonic signal gives a much clearer spatial impression than does standard mono, even without being processed by a spatialization algorithm.

For the professional market products, the real and final implementations of the different algorithms have been tested, not merely a simple demo version with limited functions and processing. The table itself can be found in Appendix D.

2.6 Other techniques for 3D sound simulation

After this overview of commercially available products for 3D surround and virtual surround systems, other state-of-the-art techniques and systems for 3D sound recording

38 *See* http://apple.com/audio/openal.html

39 *See* http://connect.creatielabs.com/openal/default.aspx

and reproduction currently under research in various centres all around the world should also be mentioned.

2.6.1 First Order Ambisonic and A-B-C-D Formats[40]

Ambisonic aims to offer a complete hierarchical approach to directional sound pickup, storage or transmission and reproduction, which is equally applicable to mono, stereo, horizontal sound reproduction, or full "periphonic" reproduction including height information (*see* also Gerzon, 1974).

Depending on the number of channels employed it is possible to represent a lesser or greater number of dimensions in the reproduced sound. A number of formats exist for signals in the Ambisonic system, as follows: the A-format for microphone pickup, the B-format for studio equipment and processing, the C-format for transmission, and the D-format for decoding and reproduction. A format known as UHJ (Universal "HJ", where HJ are the letters denoting two earlier surround sound systems) is also used for encoding multichannel Ambisonic information into two, three or four channels while retaining good mono and stereo compatibility for "non-surround" listeners.

In order to address the need for compatibility, the ambisonic system includes an encode/decode system called UHJ. UHJ contains up to four channels, depending on the transmission medium available. If all four are available, the user can decode surround-sound with height of a minimum of eight speakers. With two or three channels, the user can decode horizontal surround (three channels - one of which can be bandwidth-limited, which offers somewhat better imaging than do two) on four or more speakers. With two channels, two speakers, and no decoder at all, the result is a very enjoyable stereo - wider than the speakers and a good deal more solid, imagewise, than ordinary panpotted mixes. Sum the two channels and the result is a very good mono signal - with the difference that it provides, however, all of the punch the user would have achieved if it had been mixed in mono.[41]

40 *See* http://www.ambisonic.net

41 *See* from http://www.ambisonic.net/NewAgeAmbi.html

A B-format signal could easily be recorded using a Soundfield microphone[42] (a special microphone with four capsules arranged in a tetrahedric array; the Soundfield microphone has an A-format output, which can then be converted into a B-format signal).
To explain more simply: the B-Format works as a three dimensional extension of M/S encoding[43]. Instead of an omni and a left-right facing figure-of-eight (L+R and L-R respectively), three figure-of-eight microphones are placed at 90° angles to each other. This gives a mono signal (called W), front-back (X), left-right (Y) and up-down (Z) (*see* Figure 1).

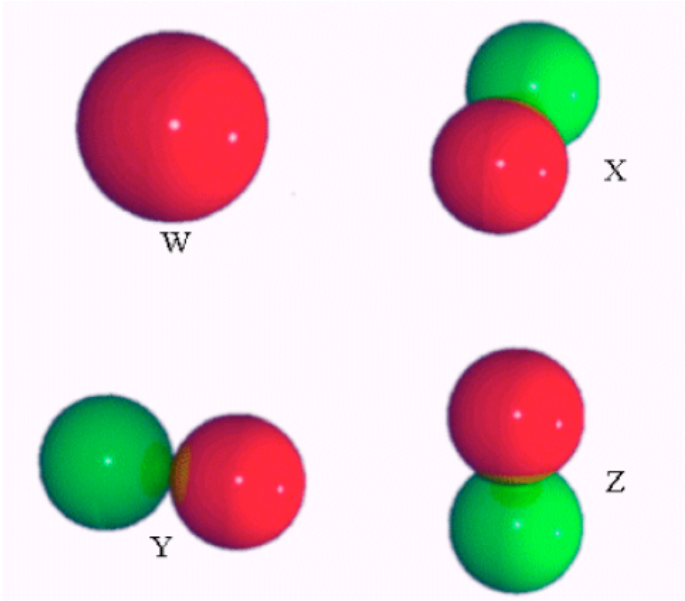

Figure 1. Directivity diagrams of the four components of the B-format signal

The interesting aspect of this encoding method is the possibility, for the reproduction of the B-format signal, of using a loudspeaker system configured according to the Ambisonic method (this is properly described as a conversion between B-format and D-format); what is achieved is to encircle a suitable volume for the listening with an adequate number of loudspeakers uniformly placed in the environment. The easiest configuration is with eight loudspeakers placed at the zeniths of a cube. From the B-format signal, it is a straightforward matter to recalculate the signal that has to be sent, for example, to the array of eight loudspeakers described. Using the loudspeaker's

42 *See* http://www.soundfield.com

43 *See* http://en.wikipedia.org/wiki/Joint_stereo

numeration in Figure 2 and the traditional conventions on the orientation of the X, Y and Z axes, and taking in consideration the fact that a Soundfield microphone produces a W signal with a gain reduced of –3dB compared to the other three channels, it is possible to obtain this group of relations to rebuild the signal which has to be sent to the eight loudspeakers:

$F1 = W + X + Y + Z$

$F2 = W - X + Y + Z$

$F3 = W - X - Y + Z$

$F4 = W + X - Y + Z$

$F5 = W + X + Y - Z$

$F6 = W - X + Y - Z$

$F7 = W - X - Y - Z$

$F8 = W + X - Y - Z$

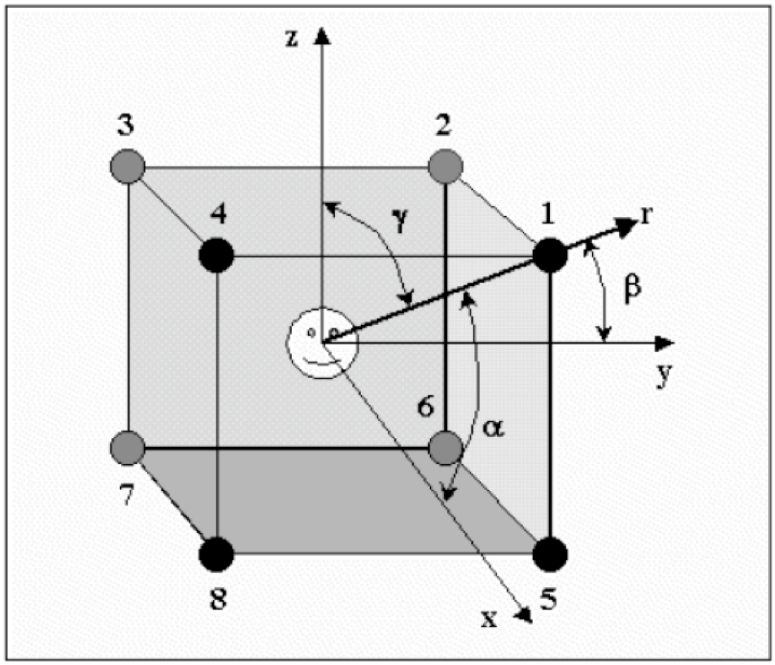

Figure 2. An Ambisonic 3D reproduction system through eight loudspeakers

As previously stated, there are other Ambisonic formats:

- A-format: consists of the four signals from a microphone with four sub-cardioids capsules orientated as in the four sides of a tetrahedron; they correspond to left-front (LF), right-front (RF), left-back (LB) and right-back (RB), although two of the capsules point upwards and two point downwards.

- C-format: consists of four signals L, R, T and Q, which conform to the UHJ hierarchy, and are the signals used for mono or stereo-compatible transmission or recording. It is a useful consumer matrix format: L is a two-channel compatible left channel, R is the corresponding right channel, T is a third channel allowing more accurate horizontal decoding, and Q is a fourth channel containing height information.
- D-format: are those distributed to loudspeakers for reproduction, and are adjusted depending on the selected loudspeaker layout. They may be derived from either B- or C- format signal using an appropriate decoder (as the matrix one shown above).

2.6.2 HOA (Higher Order Ambisonic)

A particularly active area of current research is the development of "higher orders" of Ambisonics. These use a larger number of channels than the original first-order B-Format, and offer benefits including greater localisation accuracy and better performance in large-scale replay environments such as performance spaces.

The higher orders correspond to further terms of the multipole expansion of a function on the sphere in terms of spherical harmonics. As discussed in Wave Field synthesis (WFS, Section 2.6.3), in the absence of obstacles, sound in a space over time can be described as the pressure at a plane or over a sphere – and thus if one reproduces this function, one can reproduce the sound received at a microphone at any point in the space pointing in any direction.

Michael Gerzon's original article (1974) on periphony gave versions of what was later to become known as Ambisonic encoding up to the Third Order. However, because these were given in Cartesian coordinates and all of the later work was published using polar coordinates (and with different weightings), few people have ventured into the territory above the published First Order set, at least for anything other than horizontal work. Moreover, the definitions, formulations and nomenclature of spherical harmonics show a considerable degree of variation from textbook to textbook depending on whether the field of application is in maths, physics, chemistry, or engineering (*see* Malham, 1999, and Jerome, 2000).[44]

44 *See* http://www.york.ac.uk/inst/mustech/3d_audio/secondor.html

2.6.3 WFS (Wave Field Synthesis)[45]

Wave Field Synthesis (WFS) is a spatial audio rendering technique used for the creation of virtual acoustics environments (*see* Theile, 2004). The technique is based on the production of an artificial wave front, generated from virtual sound sources, synthesized using a large number of individually driven loudspeakers. Differently from other spatialization techniques such as standard stereo, VBAP and Ambisonics, the localization of virtual sources in a WFS simulation does not depend on or change with the position of the listener. The weakness of this simulation typology is the high computational cost and the high number of loudspeakers: for this reason, WFS systems are usually limited to 2D simulations, often with only frontal and sides loudspeakers' array (no speakers on the back of the listener).

2.6.4 VBAP (Vector Base Amplitude Panning)[46]

Vector Base Amplitude Panning (VBAP) can be considered as an extension of the standard stereo panning technique applied to multi-speaker setups. In the 3D space around the listener, the presence of a sound source is simulated applying the tangent panning law between the closest triplet of loudspeakers (pair if the simulation is only in 2D). This technique is particularly simple and flexible, and can be adapted virtually to any loudspeaker system composed of more than 3 (2D) or 4 (3D) elements.

> Vector Base Amplitude Panning (VBAP) is a method for positioning virtual sources to arbitrary directions using a setup of multiple loudspeakers. In VBAP the number of loudspeakers can be arbitrary, and they can be positioned in an arbitrary 2-D or 3-D setups. VBAP produces virtual sources that are as sharp as possible with the current loudspeaker configuration and with amplitude panning methods, since it uses at one time the minimum number of loudspeakers needed, one, two, or three. (Pulkki, 1997)

2.6.5 DirAC (Directional Audio Coding)[47]

Directional Audio Coding (DirAC) is a method for spatial sound representation virtually applicable to any arbitrary audio reproduction method. The 3D audio signal is analysed, depending on frequency and time, regarding the diffuseness and the direction of arrival:

45 *See* http://www.syntheticwave.de/3D%20Wave-Field-Synthesis.htm

46 *See* http://www.acoustics.hut.fi/research/cat/vbap/

47 *See* http://www.acoustics.hut.fi/research/cat/DirAC/

this information is then transmitted, together with a mono channel for the audio content, to the decoder, which will then reproduce the 3D sound using different strategies. This technique is extremely flexible, and allows high quality 3D sound signals to be compressed and transmitted (for example for tele-conferencing applications) with a minimal loss of spatial information.

> Directional audio coding (DirAC) is a technique for various tasks in spatial sound reproduction. It is based on "Spatial impulse response rendering" [48], on the same principles, and partly on the same methods. The processing can be divided into three steps:
>
> Analysis: the sound signals are divided into frequency bands using filterbank or STFT. The diffuseness and direction of arrival of sound at each frequency band are analyzed depending on time.
>
> Transmission: a mono channel is transmitted with directional information or, in applications targeting for best quality, all recorded channels are transmitted.
>
> Synthesis: the sound at each frequency channel is first divided into diffuse and non-diffuse streams. The diffuse stream is then produced using a method that produces maximally diffuse perception of sound, and non-diffuse stream is produced with a technique which produces as point-like perception of sound source as possible.
>
> Synthesis can be implemented in various ways, depending on the microphone technique, transmission type, and reproduction system.
>
> (*See* also Merimaa, 2005 and Pulkki, 2006)

2.6.6 Stereo dipole (and other transaural systems)

> Under certain circumstances, it is possible to give a listener the impression that there is a sound source referred to as a "virtual source", at a given position in space where no real sound source exists. One way to achieve this is to ensure that the sound pressures that are reproduced at the listener's ears are the same as the sound pressures that would have been produced there by a real source at the same position as the virtual source. When only two loudspeakers are used for the reproduction, this can be achieved by using digital signal processing techniques in order to compensate for the "cross-talk". (Kirkeby, 1997a and 1997b)

When reproducing binaural stereophonic signals, it is essential that the signal from the left channel reaches only the left ear, and *vice versa* for the signal from the right channel and ear, and this is the case of sound reproduction over headphones. When reproducing

48 *See* http://www.acoustics.hut.fi/research/cat/sirr/

stereophonic signals over a pair of frontally placed loudspeakers, the signal from the left channel reaches both the left and the right ear, and the same for the signal from the right channel. This phenomenon is known as *crosstalk*. In order properly to reproduce a binaural stereophonic signal over two frontally placed loudspeakers, filters need to be used for eliminating the crosstalk effects (the function of these filters is known as *crosstalk cancellation*): the word *transaural* is used to refer to systems where such filters are implemented, and stereo dipole is one of those.

Stereo dipole is a technique for a 3-D sound generation using two loudspeakers. With the stereo dipole technique it is possible to enhance a standard stereo listening, or to reproduce a binaural recording using two frontal loudspeakers.

Instead of placing the two loudspeakers at two angles of an equilateral triangle (with the listener at the third angle), as in a standard stereo setting, the two loudspeakers are placed close to each other, at 10° of span (they could also be housed in the same cabinet).

The schematic diagram of the stereo dipole sound reproduction system is illustrated in Figure 3.

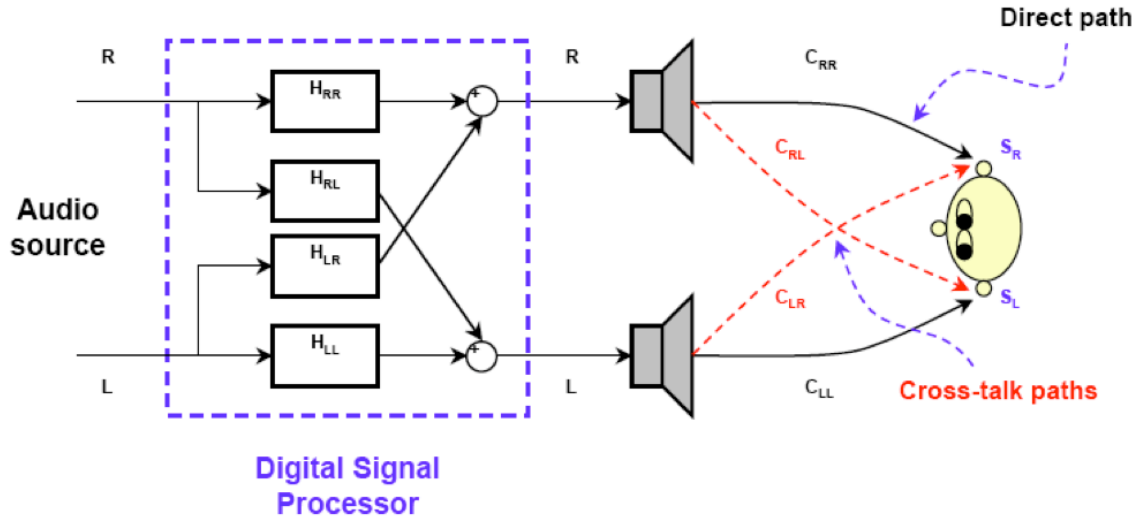

Figure 3. Configuration of a stereo dipole system

The stereo dipole is composed of a digital filter network and two loudspeakers placed very close to each other (side by side) in front of a listener.

The internal scheme of the digital filter network, is the following:

$f_{ll} = (h_{ll})$

$f_{lr} = (-h_{lr})$

$f_{rl} = (-h_{rl})$

$f_{rr} = (h_{rr})$

This is on the basis that "ll" is the signal out of the left loudspeaker that goes to the left ear, "lr" is the signal out of the left loudspeaker that goes to the right ear, "rl" is the signal out of the right loudspeaker that goes to the left ear, and "rr" is the signal out of the right loudspeaker that goes to the right ear.

When the sounds recorded by a dummy head microphone are reproduced through stereo dipole, the inverse filter that cancels the cross talk are to be incorporated as a filter matrix:

$f_{ll} = (h_{ll}) \otimes InvDen$

$f_{lr} = (-h_{lr}) \otimes InvDen$

$f_{rl} = (-h_{rl}) \otimes InvDen$

$f_{rr} = (h_{rr}) \otimes InvDen$

$InvDen = InvFilter\ (h_{ll} \otimes h_{rr} - h_{lr} \otimes h_{rl})$

2.6.7 Ambiophonics[49]

The Ambiophonics 3D sound technique is based on a mixture between Ambisonic systems (*see* Section 2.5.1), re-creating the diffuse soundfield around the listener, and stereo dipole (*see* Section 2.5.6), used as crosstalk cancellation technique for binaural signals, in order to reproduce the direct signal and early reflections.

> The Ambiophonics method combines an exploitation of seldom applied, but well documented, psychoacoustic principles with the basic rules of good musical performance space design to create believable concert-hall sound fields in dedicated home listening rooms. Ambiophonics moves the listener into the same space as the performers, by accommodating to individual external ear and head characteristics, minimizing interaural correlation at the listening positions, abandoning the traditional stereo loudspeaker equilateral triangle, recreating early reflections and reverberant fields via computer, eliminating front-loudspeaker crosstalk, and reducing the home music theater wideband reverberation time to less than .2 seconds. The completion and testing of the first full-scale version of the Ambiophonics Home Concert Hall has demonstrated that the Ambiophonic sound reproducing technique is a worthy successor to both stereophonic or surround-sound listening configurations, for staged

49 *See* http://www.ambiophonics.org

> music, in that it can consistently generate a "You Are There" concert, opera or pop sound field even preferably from standard LPs, DVDs or CDs that the ear-brain system will accept as real. (Farina, 2001)

2.7 Research Group (brief overview)

In order to complete this investigation on the state of the art in the sound spatialization field, a brief overview will be given of the different research groups around the globe that perform, or have performed, research into the binaural spatialization field. Appendix A gives an extensive table showing all of the research groups, individuals, and respective research topics and publications. The following lines offer a summary grouping them according to six different research topics (bibliographical references to those listed can therefore be found in the table in Appendix A, as well as the dates when the last information was gathered from that specific research centre).

2.7.1 Distance Perception (and Binaural Reverb)

- The Acoustic Laboratory (Aalborg University): studies distance perception, most of all in reverberant fields, with reflections of the sounds on the walls of a room; there are no studies of distance perception in anechoic fields, thus no independent analysis of the variation of IID, ITD and DDF in the function of distance.
- AMES Spatial Auditory Display (NASA): the development of an auditory display system implementing the binaural and reverberation techniques to increase the intelligibility of speech, and of different speech from different virtual directions.
- Binaural Hearing Lab (University of Boston): the development of a virtual acoustic environment for the development of capabilities for spatializing sound in simple (anechoic) and complex (reverberant) environments,
- Centre for the Neural Basis of Hearing (University of Essex): the development of physical models of the pinna and of the concha. This can be considered as a starting point for the development of a theory of distance perception, in terms of variations in localization cues.
- CIPIC Interface Laboratory (UC Davis): in the CIPIC HRIR database there are no data about distance perception itself, merely on azimuth an elevation variations; however, a study is being carried out into the individual importance of respective localization cues, and this can be considered as being particularly useful for the comprehension of the mechanism of distance perception.

- The Hearing Robot (University of Aizu): studies into the sound localization mechanism in reverberant environments for automatic recognition by a computer. This kind of algorithm implements techniques for echo cancellation, useful for the isolation of the localization cues for distance perception according to the acoustical parameters of the room where the experiments will be conducted.
- Room Acoustic Team (IRCAM): within the binaural spatialization functions in Spat, a simple binaural reverb is implemented, based mainly on the conversion of multichannel streams into binaural, considering also the directivity of the sound source to be simulated
- Keith Martin (MIT Media Lab): studies are being conducted into the cone of confusion (*see* Chapter 3), a particularly important effect that influences sound localization for close sound sources.
- The Virtual Acoustic Project (University of Southampton): research is being conducted into the simulation of ellipsoid HRTFs (a database calculated from an ellipsoid simulation of the head). The simulations are computed at different azimuths and elevations, and also at different distances (25 cm, 1 m, and 10 m).
- Immersive Audio Lab (Integrated Media Systems Center, University of Southern Carolina): research projects including skills regarding the influence of reverberation on distance perception.
- Institute of Communication Acoustics (Ruhr Universität Bochum): research into problems related to simulating reflective environments, since reflected sound has a dominant influence on the auditory spatial impression.
- LIMSI-CNRS: studies have been carried out into the ILD modification for close sound sources, and an algorithm for the simulation of these has been implemented as a MaxMSP object library.
- IEM (Graz): a whole Pure Data library has been implemented for the simulation of sound sources and environmental acoustics in the Ambisonics domain, with the option of having a binaural output.

2.7.2 HRTF Measurement or Simulation

- AMES Spatial Auditory Display (NASA): research has been done into the measurement of a HRIR database using Golay code sequences, and implementing a

technique for the isolation of the reflection from the direct sound (thus the elimination of the need for an anechoic chamber). Another technique is implemented for a simplified HRTF FIR filter.

- AUDIS catalogue of human HRTFs: HRIR database from 12 individuals.
- Binaural Hearing Lab (University of Boston): research into a virtual acoustic environment, in order to simulate an "empirical HRTF".
- CIPIC Interface Laboratory (University of California Davis): HRIR database measured from a dummy head and various individuals.
- HDRL – Hearing Development Research (University of Wisconsiin-Madison): measurement of an HRIR database (information about the head used for the measurement was not available).
- MIT Media Lab: HRIR measurements of a KEMAR dummy head microphone.
- IRCAM Room Acoustic Team: Listen HRTF, measurement of a HRIR database from various subjects and from a dummy head.
- The Virtual Acoustics Project (University of Southampton): HRIR measurements using MLSSA (Maximum Length Sequence System Analyzer) and MLS signal technique.

2.7.3 HRIR Interpolation Techniques (or other techniques for the simulation of sound source movements)

- Binaural Hearing Lab (University of Boston): research into a virtual acoustic environment, with functions for HRIR interpolation.
- CIPIC Interface Laboratory (University of California Davis): MTB, Motion Tracker Binaural Sound, a system for the recording, the storage, and the reproduction of binaural sound, with head motion tracked functions (changing of the soundscape following the movements of the head). This does not implement a HRIR interpolation technique, although the result can be considered similar.
- CSULA Psychoacoustic Web Page (California State University, Los Angeles): research projects on the Auditory Motion Discrimination (correlations with MAMA, Minimum Audible Movement Angle).
- IRCAM Room Acoustic Team: implemented inside the SPAT MaxMSP library, there is an interpolation technique, based on the separation of ITD and DDF (*see* Section 3.4).

- LIMSI-CNRS: a MaxMSP implementation of a HRIR interpolation technique based on the independent management of ITD has been carried out.
- IEM (Graz): sound source movements are calculated in the Ambisonics domain, then converted into the binaural domain.
- The Virtual Acoustics Project (University of Southampton): visually adaptive imaging, a selection of appropriate virtual audio filters (with interpolation techniques between these) corresponding to a listener's varying head positions.

2.7.4 HRTF Quality Testing

- The Acoustic Laboratory (Aalborg University): the development of a method for the determination of the hearing threshold, monaurally and binaurally.
- Acoustic Research Centre (University of Salford): the development of a measurement method for comparing spatial performance of different reproduction systems, and on the "Clean Audio Project", for the determination of the quality of the surround sound.
- Binaural Hearing Lab (University of Boston): the development of signal processing models for the calculation of the performance of psychophysical tasks at different levels of the auditory pathway.
- Centre for the Neural Basis of Hearing (University of Essex): the development of methods for the absolute judgement of the location of auditory sound sources (using mannequin recording, with a KEMAR dummy head); four different experiments are carried out in order to assess the accuracy of the localization of auditory sound sources.
- CIPIC Interface Laboratory (University of California Davis): the development of a methodology towards understanding the relative importance of the localization cues (ITD, IID, DDF).
- HUT Acoustic Lab (Helsinki University of Technology): the development of a methodology for the psychoacoustic evaluation of spatial sound through listening tests.
- William Martens spatial hearing research (McGill University): perceptual evaluations of HRTF filtering.

2.7.5 Human Ear, Head and Auditory System Physical Models

- The Acoustic Laboratory (Aalborg University): research into the description of sound transmission in the ear canal. These results are important for the programming of a headphone linearization filter (in fact, in this research group work has been done into a methodology for the design and measuring of headphones).
- Binaural Hearing Lab (University of Boston): the development of a model of brainstem and midbrain neurons; the focus of this research is into how the neurons respond to the ITD and ILD.
- Centre for the Neural Basis of Hearing (University of Essex): pinna models, and physical models of the human concha.
- CIPIC Interface Laboratory (University of California Davis): the development of HRTF models and model composition.
- R. O. Duda Research: the development of computational models of the human sound localization process.
- HUT Acoustic Lab (Helsinki University of Technology): binaural auditory modelling.
- Ewan Macpherson's Research (Central System Laboratory, University of Michigan): the development of models for the binaural localization system.
- William Martens spatial hearing research (McGill University): HRTF simulations and generalized HRTF.
- Keith Martin's Research Interests (MIT Media Lab): spatial hearing model.
- Parmly Hearing Institute (Loyola University of Chicago): neural modelling, auditory image models.
- Sheffield Speech and Hearing Research Group (University of Sheffield): research into incorporating binaural cues in a computational model of auditory scene analysis.
- Spatial Audio Work (Georgia Institute of Technology): environmental modelling for binaural audio.
- The Virtual Acoustics Project (University of Southampton): BEM simulation of ellipsoid HRTF (using a Matlab ellipsoid model).

2.7.6 Spatial Hearing and Vision

- AMES Spatial Auditory Display (NASA): research into auditory and visual displays.
- Centre for the Neural Basis of Hearing (University of Essex): sound localization in VR systems.
- CSULA Psychoacoustic Web Page (California State University, Los Angeles): relationships between the auditory and visual systems. Studies are being performed into visual search, visual/auditory dominance and concurrent sound localization.
- DIVA – Digital Interactive Virtual Acoustic Telecommunication (HUT, Helsinki): EVE – Experimental Virtual Environment, started from a project of a virtual orchestra, with 3D audio technology through headphones synchronized with a video virtual reality.
- Human Interface Technology Laboratory (University of Washington): the development of VRD, a three-dimensional display technology for US Navy Pilots.

2.8 Final considerations

At the end of this state of the art research into the sound spatialization and binaural spatialization fields, the following summaries can now be made:

- Localization of the apparent image of the sound sources outside the head: in none of the binaural spatialization algorithms that have been tried was it possible properly and precisely to perceive sound sources located outside the head. Even if sometimes the soundscape seemed to move from inside to outside the head of the listener, the positioning of the sound sources was everything but clear and precise. Given the work of the different research groups, a few have indeed been specifically focusing on this issue, also known as the "externalization of sound sources". AMES (NASA) has carried out extensive research into this topic, but unfortunately only very little information is available on their research. Other publications have been written by various researchers (*see* Thomas, 1997; Weinrich, 1992; Hartmann, 1996; Takeuchi, 1998, and Brookes, 2005), yet no actual software or hardware implementations are available.
- Simulation of distance: a few algorithms implement the simulation of distance. Panorama (WaveArts) is the one with the largest number of control parameters such

as reverb, or reflections, but the simulation is done only through the loudness and the direct/reflected sound ratio. No systems could be found with simulation of the spectral cues for distance perception. Considering the work of the different research groups, only a few have focused on the variation of localization cues in the function of distance; much research concentrates more on distance perception in reverberant fields, therefore with the influence of all of the acoustical parameters of the room, than on the individual analysis of the localization cues for distance perception. An HRIR database with distance parameter variations was not found; all of the most important results (MIT and CIPIC) are measured at a fixed distance from the head.

- Binaural reverb: it was difficult to find systems that implemented reverberation in the binaural domain. Panorama (WaveArts) is the one that offers more parameters in terms of reverberation, although from the information that could be gathered it is a simple binaurally enhanced stereo reverb. The Beyerdynamic binaural simulation indeed offers a reverb, with its generation limited to a 5.1 setup. Considering the work of the different research groups, both IRCAM (SPAT) and the IEM (Graz) developed binaural reverbs based on multichannel (mainly Ambisonics) rendering: very few binaural reverb algorithms based on BRIR (Binaural Room Impulse Response) could be found, and for these the flexibility of the environmental simulation was simply limited to the room in which the BRIR was measured.
- Localization of frontal sound sources: in all of the reviewed binaural simulations, it was particularly difficult to perceive apparent sound sources located in the front hemisphere. This is undoubtedly a common problem for all the binaural spatialization algorithms. Few researchers have been working on this topic (*see* Weinrich, 1992 and Thomas 1997), but, as stated above, for the "externalization of sound sources" issue, no implementations are available.
- Multichannel to binaural conversion: even if algorithms and systems for the conversion between 5.1 streams and binaural were present on the market, it was not possible to find any easily accessible system that could convert offline any multichannel stream (given the position of every single loudspeaker to be simulated) into a binaural one, performing also environmental simulation (the IEM Bin_Ambi library allows Ambisonic to binaural conversion, but it is far from being "easily accessible" to individuals lacking proficiency skills in using Pure Data).

- HRIR Database: several HRIR databases are actually being used in the different algorithms and systems listed in this chapter. Indeed, the most ubiquitous is the MIT Kemar HRTF, but this does not mean that it can any longer be considered a high quality HRIR database (measurement was in fact done in 1994; *see* Gardner, 1994). Some comments can be made:
  - Measurement signal: most of the HRIR databases (included the MIT one) have been measured using the Golay Code signals, or the MLS signal. Both techniques can stimulate different kinds of errors during the measurement (*see* Picinali, 2006, and also Chapter 4 of this thesis). It has been demonstrated that the best signal to be used for these applications is a sine sweep (*see* Muller, 2001), and only the IRCAM Listen database has been measured thus (an evaluation carried out on the material is actually available online).
  - Sample Rate and Bitrate: the MIT database has been measured at 44.100 kHz and 16 bits. The IRCAM Listen database has been measured at the same sample rate, but at 24 bits of resolution: higher sample rate databases are not available at the moment.

It can clearly be noted that, nowadays, algorithms and systems for the binaural spatialization are far from being 100 per cent effective. There exists room for improving the knowledge already produced, and for creating a complete binaural tool to assist in resolving the issues outlined in this chapter.

In the introduction to this thesis, a subsection can be found named *Contributions of this research to the state of the art*, in which this research is put into the context of all of the information gathered in this chapter about the consumer and professional market, about the other research groups and researchers working on this topic, and about establishing the innovative characteristics and functions of the "binaural tool" developed in this specific research work.

Chapter 3

3. Binaural Phenomena for the Perception of the Angle

How can the human hearing system determine the direction of the sound it perceives? In this chapter the mechanisms of spatial hearing related to the perception of the angle of incidence will be described and analysed, starting from the interaural differences and arriving at the monaural cues, reviewing then also specific psychoacoustic effects linked with binaural perception.

No innovations are presented within this chapter; this is simply an introduction to the binaural phenomena for angle perception (much as Chapter 5 will be an introduction to the binaural phenomena for distance perception); therefore theories presented within the chapter have already been introduced in the past by other researchers (references can be found within the chapter and in Blauert, 1996, Moore, 2003 and Yost, 2000). However, this introduction is essential for the reader in understanding that which will be explained in Chapter 4.

A similar introduction to these phenomena has already been published by the author (Picinali, 2009). In this case, too, it was merely an introduction, and the original parts presented were related to the second part.

3.1 The Localization Cues

As defined in Chapter 1, sound localization is the judgement made regarding the specific location of a sound source, and it is performed through particular mechanisms fulfilled by the human auditory system. In order to accomplish these mechanisms, the auditory system can work on certain particular attributes of the signal input into the ear canal. Those are known as the localization cues; they can further be distinguished into interaural differences and monaural attributes, as will be shown in the following sections.

3.2 The Interaural Differences

There are two kinds of interaural differences for the localization of sound sources at the left or at the right of the head:

- ILDs: Interaural Level Differences, the differences in terms of the pressure level of a sound stimulus between one ear and the other. The differences are generated by the presence of the head between the ears, the head that acts as an obstacle placed in a direct path between the sound source and the contralateral ear entrance.
- ITDs: Interaural Time Differences are the differences in terms of the arrival times of a sound stimulus between one ear and the other. The differences stem from the different paths the sound wave needs to cover in order to arrive from the sound source to each ear respectively. When the sound source is not located in the median plane, the distances between it and each ear will differ; therefore, the sound wave will take a relatively longer or a shorter time to reach the beginning of each of the ear canals.

Both perceptions are effective in order to perform the localization of a sound source placed on the left or on the right of the head; however, their importance varies according to the frequency bands covered by the sound source to be localized. The *duplex theory* (*see* Lord Rayleigh, 1907) is probably the most widely accepted hypothesis as to how the interaural mechanism works for sound source localization. The theory shows that ILDs are more effective for the localization of high frequencies, as the ITDs are for low ones.

To clarify: as is known, low frequencies (between 20 Hz and 500 Hz, with wavelengths ranging from 16 m to 64 cm) have a wavelength that is much larger than the diametre of the head (~17 cm); therefore, the low frequencies will not be scattered or absorbed by such a small obstacle. Thus, the ILDs are found to be almost irrelevant for low frequencies, yet significant for high frequencies. As an example, after Figure 1, it can be noticed that when a sound source is located at 90° of azimuth, the ILDs are at 1-2 dB for the 200 Hz frequencies, and at 20 dB for the 6000 Hz. For this reason, the ILDs can be viewed as a frequency-dependent parameter.

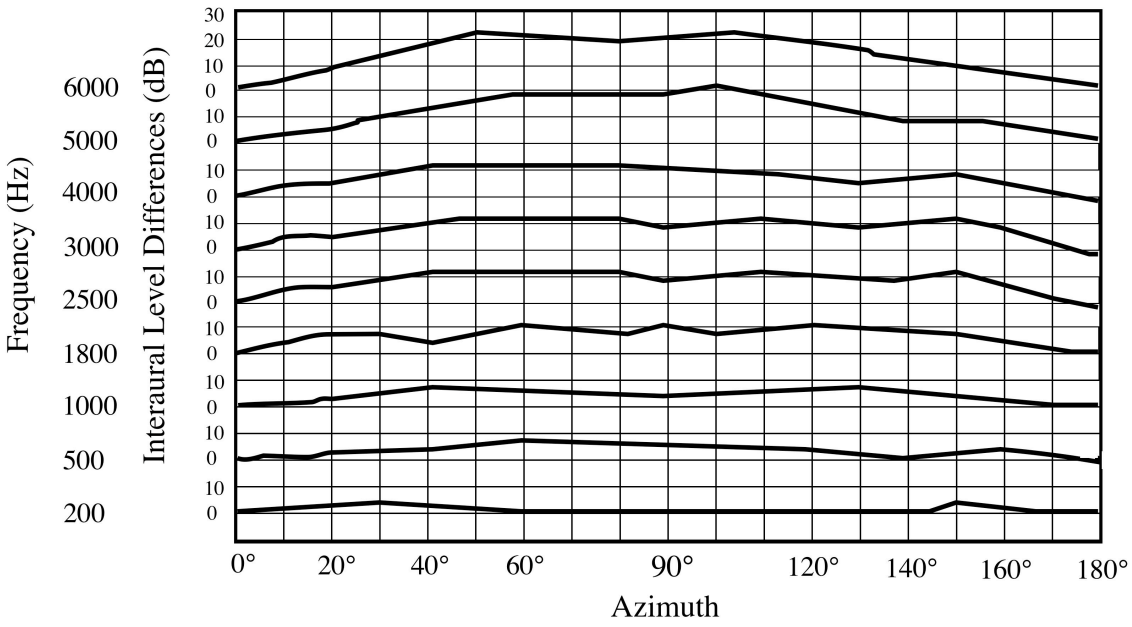

Figure 1. The interaural level differences in frequency bands as a function of the azimuth (after Feddersen, 1957, redrawn from Moore, 2003)

Figure 2, below, shows how the ITDs vary independently of the frequency of the stimulus; in fact, the time taken by a sound wave to travel from the sound source to the two ears is dependent not upon the frequency, but upon other physical parameters such as air temperature, pressure and humidity, which determine variations in the speed of sound. Thus, the ITDs can be seen as a frequency-independent parameter.

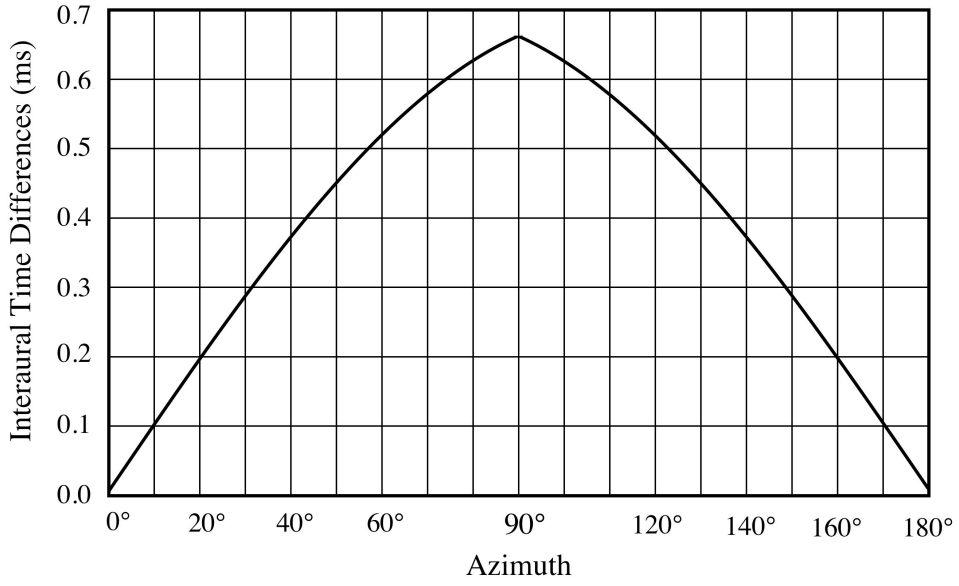

Figure 2. The Interaural Time Differences as a function of the azimuth (after Feddersen, 1957, redrawn from Moore, 2003)

However, the fact that the ITDs can be measured in terms of microseconds (a millionth of a second) creates some detection problems for frequencies whose period is comparable with that of the ITDs themselves. For example, if a sound source generating a 2000 Hz pure tone were located at 60° of azimuth, the ITDs (calculated for a subject with a head circumference of 58 cm) would be 0.5 milliseconds, exactly the period of the 2000 Hz frequency. In this case, it would be utterly impossible to establish the position of the sound source using only the ITDs as a determinant, because the sound waves at both ears would be in exactly the same phase, and would not be distinguishable except for the 0.5 milliseconds difference in the onset of the oscillations. These problems are magnified for higher frequencies and smaller periods.

In this scenario, the basis of the *duplex theory* may be understood: for certain frequencies, the ILDs seem to be the more reliable parameter for left-right localization, while for others the ITDs can be considered more effective. It has been calculated that for frequencies above 725 Hz (with periods shorter than 1.38 ms, when the ITDs would start to create similar problems) the sound localization in terms of left-right detection is accomplished through considering mainly the ILDs, while for frequencies below 725 Hz (with wavelengths greater than 44 cm, when the ILDs would start to become irrelevant) the ITDs constitute the most important parameter.

It is important to underline that the presence of complex spectra and transients within signals can highly facilitate the mechanisms of sound localization (see Moore, 2003): nevertheless, this does not happen in the case of pure tones with a constant amplitude in time.

3.3 ITD vs ILD

As stated in the previous sections, when two identical stimuli are presented through a pair of headphones at the two ears, the sound source is usually lateralized inside the head of the listener, more or less in the middle of the line between the two ears. If, instead, one of these two stimuli arrives at the right ear with a delay of, for example, 100 microseconds, the sound image is moved towards the left of the listeners' head, and *vice versa* if the delay is applied to the left ear. It is nevertheless possible to counterbalance this difference through introducing an interaural level difference (ILD) between

the stimuli at the two ears; the time difference needed to "centre" an ILD of 1 dB is called the *Trading Difference* (also *Trading Ratio* or *Compensation Factor*).

In one of the first studies on this topic, a theory has been elaborated (*see* Deatherage, 1959) which embraces the idea that the time and level differences are coded in the same way by the neural hearing system (for an explanation of the internal hearing system, *see* Chapter 1): a high intensity stimulus causes a quicker neural response that does a low intensity one, and in this way the ILDs are basically transformed into ITDs (also known as the *Latency Hypothesis*, *see* Blauert, 1996:167).

Nevertheless, it has more recently been demonstrated that the human hearing system does not work in exactly this way (*see* Whitworth, 1961 and Hafter, 1968; 1972). It has been in fact proved that, both for low frequency tones and for clicks, when ITDs are used to compensate ILDs, the perception may exist of two distinct sound sources (for a more "artistic" approach to this problem, *see* Picinali, 2009). These experiments confirm the fact that ILDs and ITDs are not equivalent; they are not perceived in the same way by the human hearing system, and are therefore not interchangeable.

3.4 DDF (Direction Dependent Filtering)

3.4.1 The Cone of Confusion

Thus far, it has been explained how it is possible to determine the provenance of a sound from left and from right on the horizontal plane, while the issues linked with the differentiation of sound sources placed in front of or behind the listener, as well as above or below, have not yet been addressed. If two sound sources are located at 60° and 120° of azimuth (two positions that are specular, in the context of the frontal plane), the interaural differences in the signals coming from them will be exactly the same, even if the two sources have different real positions. This problem is called the *Cone of Confusion* because plotting all of the positions of the sound sources with the same interaural differences would generate the shape of a cone, with the head position as the apex (*see* Figure 3), and it can be resolved merely with the help of a third localization cue, the direction-dependent filtering.

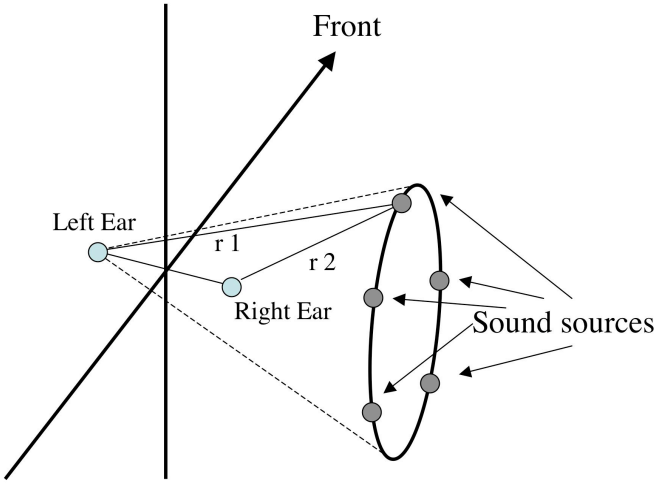

Figure 3. The Cone of Confusion. For each of the sound sources located in the circle (which can be considered as the base of a cone), r1 and r2 are respectively equal, therefore the interaural differences are exactly the same (von Hornbostel, 1920, redrawn from Blauert, 1996)

3.4.2 The Direction-Dependent Filtering and the Head Related Transfer Function

Chapter 1 described the two main functions of the pinna; while the first, sound gathering, can readily be understood, the second, direction-dependent filtering, appears more complex. The dimensions of the pinna are far too small, if compared with the wavelengths of many audible frequencies, for it to function as a simple sound reflector: the dimensions of its cavities, instead, are comparable to $\lambda/4$ (where λ is the wavelength of a given frequency) of a large number of frequencies, and these can easily become sound resonators for sound waves coming from specific directions. Therefore, inside the pinna the sound is modified by reflections, refractions, interferences, and resonances activated for specific frequencies and, most significantly, for the incident angles of specific sound waves, hence the name *Direction-Dependent Filtering* (*see* Batteau, 1967; 1968 and Shaw, 1968).

The ensemble of the filtering effect generated by the pinna, as well as by other elements such as the head, the torso, and the shoulders, composes the so-called HRTF, Head Related Transfer Function (*see* also Chapter 1, Basic Notions).

In order to abstract and simplify the principle, an empty bottle serves as an example: when blowing with the mouth close to the neck of the bottle, it is possible to generate a

resonance the frequency of which is determined only by the volume of the air inside the bottle and the dimensions of the bottle neck and not, for example, by the material of the bottle itself. However, in order to generate resonance, the position of the mouth, the force of the air blown, and the inclination of the bottle need to be specifically selected (a choice that is usually achieved by trying different positions and speeds). It must be asked what would happen using a 'special' bottle, with more necks, more openings, and more cavities? There would then be many more resonances, and many more combinations of positions and blowing speeds in order to activate them.

This is what happens if the pinna is considered as a complex resonator: multiple resonances can be activated depending on the incidence of the sound wave. Of course, then, the phenomena of the DDF are far more complex and, as previously stated, are generated and modified considering not only resonances, but also reflections, shadowing, dispersion, diffraction, and interferences.

In the nineteen-sixties, Batteau (*see* Batteau, 1967; 1968) tried to study the reflections of the sound waves on the pinna, to extrapolate the ratio between the direct and reflected sound at the entrance of the ear canal. Figure 4 shows his analytical model, intended to explain the effect of the pinna on the signals coming from different directions.

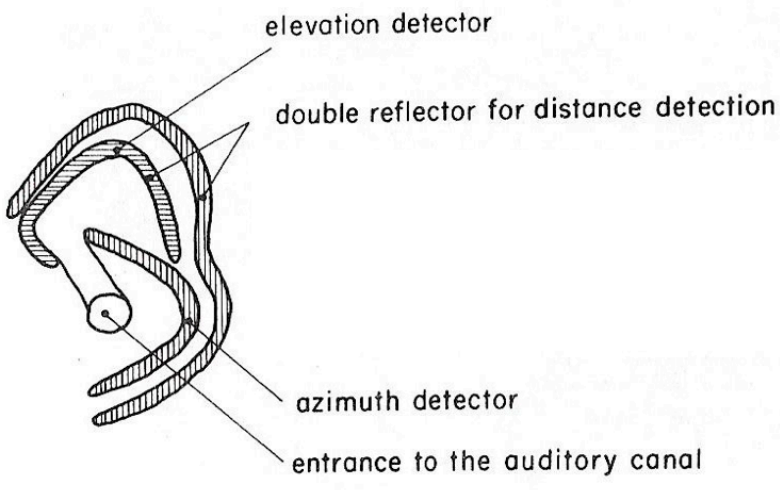

Figure 4. An analytical model intended to explain the effect of the pinna (Batteau, 1967, redrawn from Blauert, 1996)

Batteau supports the thesis that there are two distinct delay lines (plus, obviously, the direct signal) allowing the hearing system to establish the position of a sound source for

both the azimuth and elevation angle, Nevertheless, this presupposes a large amount of approximation, given the fact that the wavelengths of most of the frequencies arriving inside the pinna are far too long to allow a simple reflective phenomenon. Furthermore, it is rather imprecise to contemplate only two distinct delay lines, considering the fact that the phenomenon of the reflections inside the pinna is far more complex.

In the same period, Shaw and Teranishi (*see* Shaw, 1968) used a rubber pinna model, a probe microphone and a point sound source to measure the transfer function of the external ear in a wide variety of conditions of the incident sound. The probe microphone was positioned at varying distances from the eardrum, and the sound source was the opening of a 1 cm diametre pipe positioned at 8 cm from the entrance of the ear canal. A series of resonance frequencies were then detected and correlated with the dimensions of the cavities of the pinna itself (*see* Figure 5):

- F1 (approximately 3 kHz): it is a $\lambda/4$ resonance of a tube closed at one side, with a length of 30 mm, therefore approximately 33 per cent more than the real length of the ear canal (in this case, the pinna seems therefore to act as a prolongation of the ear canal itself).
- F2 (approximately 5 kHz): the maximum pressure of this oscillation entirely fills the ear canal; the distribution of the pressure is therefore the same as that with the eardrum occluded. The ear canal and the cavum conchae (*see* Section 1.3) are involved in this resonance, which can be modified in frequency through inserting putty inside the concha, and does not depend strictly on the incidence angle of the signal.
- F3 (approximately 9 kHz), F4 (approximately 11 kHz) and F5 (approximately 13 kHz): they are stationary longitudinal waves of $\lambda/2$ and λ, which involve the ear canal and the concha.

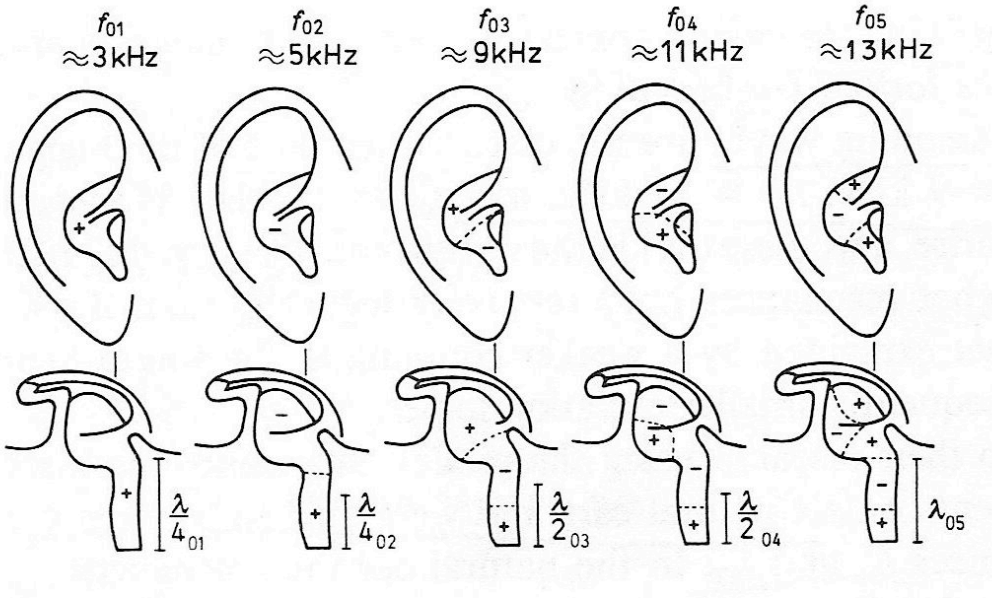

Figure 5. Distribution of the sound pressure, for different resonance typologies, inside an external ear model with a high impedance end. The dotted lines indicate the nodal points of the pressure distribution (Shaw, 1968, redrawn from Blauert, 1996)

All of these resonances can vary depending on the characteristics of the pinna of that specific individual (*see* also Section 3.4.3) and, most of all, on the angle of incidence of the stimulus (except, as mentioned above, for F2): it has in fact been noted that the peaks in normal HRTF correspond, with a good approximation, to these frequencies. The reason for this can be explained through the complex interference effect between the different parts of the pinna (as well as through the different refraction and diffraction phenomena). For example, the existence of a link has been proved between resonances at frequencies around 8 kHz and the elevation angle at which the sound source is located.

Results of experiments conducted by Blauert in 1967 (*see* Blauert. 1996) provide further clues to the details of the effects outlined by Shaw and Teranishi. Blauert's experiments were very similar to those described previously, differing only in the length of the artificial ear canal, and also outlined the link between the resonances and the angle of incidence of the sound. For example, the F2 resonance frequency remains more or less constant until 90° of azimuth, then decreases by 15-20 dB between 90° and 110°, and remains constant until 180°, demonstrating therefore that the resonance is not activated for sound sources coming from the back. Through these experiments, Blauert was also

able to demonstrate that the resonances inside the ear canal are independent of the azimuth and elevation variations, whereas the resonances that include also the cavum conchae vary. Here, for the 10 kHz frequency, at angles of 0° and 180° of azimuth, the stationary wave inside the ear canal remains unvaried, but while for 180° resonances are present also inside the cavum conchae they disappear at the end of the ear canal at an angle of incidence of zero degrees.

All of the resonances generated by the reflections and refractions on the shoulders and on the torso of the listener need to be considered, too, as outlined at the beginning of this section. It may therefore be easily understood how sound input into the auditory canal is modified through complex frequency filters that change their shapes depending on the position of the sound source, hence the name Direction Dependent Filtering.

3.4.3 Individual and general attributes of the HRTF

While some of the HRTF parameters may be considered constant for all human hearing systems, certain others need to be considered individually, because they depend on idiosyncratic physiological differences between human beings, such as the shape of the pinna and the circumference of the head. Both individual and general attributes should be considered when an HRTF is simulated, or binaural recordings are performed (*see* Chapter 4), thus a simulation should be performed individually for each subject. For further information on this topic, *see* Møller (1996) and Katz (1996).

The current research determined that no HRTF individualization would be performed; for a justification on this choice, see the introduction of the thesis.

3.4.4 The role of the head movements

It is essential to underline that head movements are extremely important to sound source localization, and most of all for the front-back and up-down discriminations, when the cone of confusion issue needs to be resolved. In fact, turning the head left-right or up-down causes important changes in the soundscape and in the relative positions of the sound sources, generating further information that may be considered as particularly relevant for the correct localization of the sound source. As an example, if a sound source is located in the horizontal plane at 60º of azimuth, it could be easily confused with another located in the same plane at 120º of azimuth; on rotating the head to the left, if the sound source is really located at 60º of azimuth it will move towards the front

(0º of azimuth); otherwise, if it is located at 120º of azimuth, it would move towards the left (90º of azimuth). The movements of the head are essential for a proper sound source localization, for example, in critical situations such as those with narrowband stimuli (sound stimuli with a narrow frequency extension, which would be less adversely affected by a filtering process on the whole frequency scale such as the DDF) or in the case of a localization task in a particularly reverberant environment, when the localization cues cannot be precisely analysed, Similarly to the statements in the previous section (Section 3.4.3), in this research work no head-tracking functions for the rotations of the simulated binaural soundfield have been implemented; for a justification of this choice, see the introduction to the thesis.

3.5 Sound localization on the three planes

After this brief overview on interaural differences and direction-dependent filtering, it should now be understood how sound source localization is performed for a source placed in a three-dimensional soundscape. The mechanisms for the localization of the sound sources when these are placed in just one plane will now be addressed, in order to simplify the circumstances.

3.5.1 Sound source localization in the horizontal plane

In this specific case, the presence of dichotic stimuli is highly probable. Solely for sound sources located exactly at 0º and 180º of azimuth (and, of course, 0º of elevation, simply because it is in the horizontal plane) a diotic stimulus can be input into the hearing system. Thus, for left-right discrimination (0º/180º to 180º/360º) the ITDs and ILDs are used, while for front-back judgement (90º/270º to 270º/90º) the DDF carries major importance. Because all three localization cues can be used and a sound source would most frequently be located here (taking speech as an example), the horizontal is the plane where sound source localization performances are higher (*see* Section 3.5.4).

3.5.2 Sound sources localization in the median plane

In the horizontal plane dichotic stimuli are far more common, whereas if a sound source is located in the median plane the sound will certainly reach the ear as a diotic stimulus. If the asymmetries of the head are ignored, the distances and the angles of incidence between a sound source located in this plane and both ears are always the same. There-

fore, the only parameter applicable to the hearing system is the DDF. For these reasons, the median is the plane with which the human hearing system has greater difficulties in terms of sound source localization performance.

3.5.3 Sound sources localization in the vertical or frontal plane

The frontal plane can be considered as the 'vertical version' of the horizontal plane. Diotic stimuli are present only for sound sources located at 90º and 270º of elevation, while for all other positions a dichotic stimulus would reach the hearing system. Up-down discriminations are performed through analyzing the DDF. The localization accuracy on this plane is not as precise as it is in the horizontal plane, while certainly not as vague as for the median plane.

3.5.4 The Minimum Audible Angle (MAA)

The Minimum Audible Angle (MAA) is the barely-perceptible difference of localization on the horizontal plane (azimuth) detectable by the hearing system. It depends on various elements linked with individual subjective factors, plus the frequency and position of the source, although under ideal conditions it may be estimated. Figure 6 shows the diagram of the MAA for sinusoidal signals varying by frequency and by azimuth.

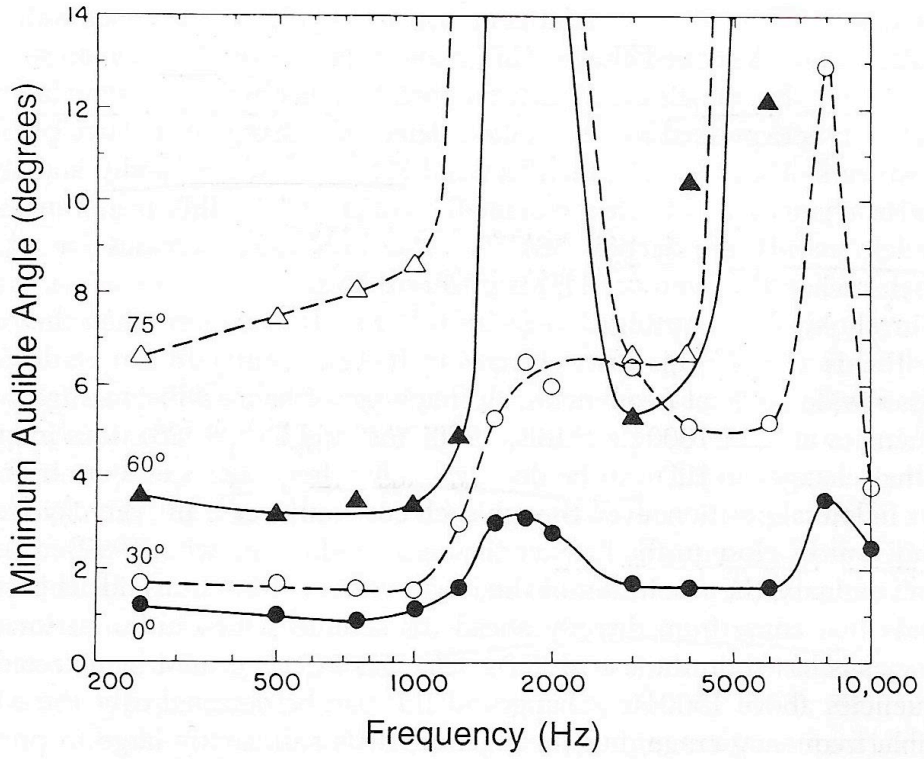

Figure 6. The MAA for sinusoidal signals varying azimuth and frequency (Mills, 1958 ; 1972, redrawn from Moore, 2003)

The optimum localization precision is demonstrated as being for 0° of azimuth, where the hearing system can discriminate angles of a little more than 1°, while the performance becomes weaker for sound source positions tending towards 90° of azimuth. The MAA also varies depending on the frequency of the stimulus: for lower frequencies, smaller angles are detectable, while above 1500 Hz the precision becomes more limited (in inverse proportion to the MAA). This feature is linked to mechanisms of the duplex theory cited in Section 3.2, for which the ITDs become non-influential on frequencies above 750 Hz.

Studies for the estimations of the barely-perceptible differences according to the angle of elevation have also been carried out (*see* Perrot, 1990, and Grantham, 2003). They showed that the hearing system localization precision for sources located on the vertical and median planes is weaker (*see* also Sections 3.2 and 3.5.1, 3.5.2 and 3.5.3), as may be seen in the diagram in Figure 7.

The importance of the MAA and of the barely-perceptible differences according to the angle of elevation is particularly significant when planning experiments for the measurement of HRTF (*see* Chapter 4). In order to have the greatest accuracy for the

binaural simulation to be performed, the grid on which the HRIR are measured needs to be established following the localization precision of the hearing system for the different positions and angles (usually, HRIRs are measured each 5° of azimuth and 10° of elevation).

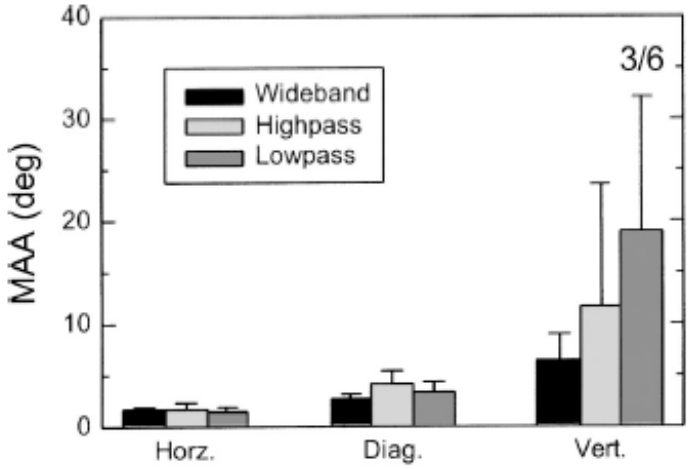

Figure 7. MAA thresholds across six subjects for the three planes of presentation: horizontal (0°), diagonal (60°) and vertical (90°). The different shadings represent the different filter conditions. Error bars indicate one standard deviation. The fraction shown over the right-most bar indicates that in this case only three of the six subjects could obtain a threshold with the low-pass signal on the vertical plane (from Grantham, 2003)

3.5.5 Localization of pure and complex tones

"Complex tones" signifies those signals that are not sinusoidal or pure; therefore, they are tones with harmonic or in-harmonic spectra that vary in intensity and frequency content in the function of time. Those tones are not affected by the phase ambiguities that create various problems for the localization of pure tones (see Sections 3.1 and 3.5.1). For signals with a broadband spectrum, ambiguities are resolved by the hearing system through its performing complex comparisons between the different localization cues measured at the different frequencies of the spectrum of the stimulus itself, allowing, for example, the determination and effective use of both ITDs and ILDs for the same signal.

For complex tones, therefore, the localization mechanisms are far more complex, while the localization accuracy become greater for all the three planes.

3.6 Binaural effects

The following section comments on three different psychoacoustic effects linked with binaural hearing. At the end of each sub-section, an explanation will be given as to the reason why that specific effect has been discussed within this thesis, and how its description will be important for the subsequent chapters.

3.6.1 Binaural beats

This is a particular phenomenon linked with the ability of the human hearing system to process small phase differences between the ears. When a signal arrives at each ear with a small interaural frequency difference (between 0 and 10 Hz, depending on the frequency of the stimulus), the tone seems to move from left to right. When the frequency difference is larger, between 10 and 20 Hz, the spatial sensation is even more effective, and fluctuations in the intensity of the signals are also perceivable. Ideally, in order to have sufficiently clear binaural separation, the signals need to be played back through headphones. This phenomenon may be considered similar to one of the beats between two similar frequencies, although it presents many differences. The beats are a physical phenomenon, while the binaural beats are a psychoacoustic phenomenon with its origins in the superior olive, therefore in the first place where the auditory nerves leading from the ears meet before arriving at the brain. Furthermore, the binaural beats phenomenon seems to be highly subjective, and therefore its perception varies considerably among individuals, although certainly more effective for lower frequencies (below 1000 Hz).
Chapter 2 lists and describes different "spatial enhancement" algorithms on the consumer market. A certain number of these were actually based upon the binaural beat effects, giving a full spatial impression from a stereo signal played back through headphones using a psychoacoustic effect based on interaural frequency differences.

3.6.2 Binaural masking and Cocktail Party Effect

Given their complexity, these two well-known binaural effects deserve a whole chapter on their own. The following lines, though, provide a brief description, and attempt to outline of what they consist and how they work in terms of spatial hearing cues.
Consider the following situation: a noise signal and a pure tone are sent diotically to both the right and left ear. The level of the pure tone is calibrated such that it is fully

masked by the noise and is thus inaudible. The level in dB of the tone will be given the name L_0. If the phase of the pure tone is inverted for one ear only, the tone will become audible again. Reducing the amplitude of the tone until it is again fully masked by the noise, this second level in dB will be given the name L_1.

The difference between L_0 and L_1 is called BMLD (Binaural Masking Level Difference), and it can go from 1 dB for low frequencies (around 300 Hz) to 2-3 dB for higher ones (above 1500 Hz, *see* Durlach, 1978). Maintaining unvaried the phase of the pure tone (the same for each ear) while inverting the phase of the masking noise will achieve the same results.

This is the so-called "binaural masking effect"; when using speech instead of a pure tone, and inverting the phase of either the masking noise or the speech signal for one ear only, the effect would be of an increase in the intelligibility of the speech. A complete phase inversion for a signal in one ear and not in the other is a situation that, self-evidently, cannot happen in a real-life situation. Having said that, a small interaural phase difference or low interaural correlation in either the target signal or the masking noise would cause an increase in the intelligibility of the target signal itself, and this brings the discussion to the description of the second effect illustrated in this section.

It may and does occur in everyday life to find oneself in a situation where more people are talking at the same level in one single room. The listener is nevertheless able to concentrate his/her attention on one single voice, isolating it from the others, even without rotating the listener's head towards the speaker. If the listener uses then a hand to close one of his/her ears, the ability to isolate and understand that single speaker becomes much weaker, therefore the intelligibility of the speech decreases significantly. This phenomenon is known as the Cocktail Party Effect (*see* Cherry, 1953); it is based on the fact that a particular signal S with a given incidence direction is less masked by a noise N, coming from a different position, if the subject is listening binaurally rather than monaurally (as if with one ear closed).

The reason why these two binaural effects have been cited and briefly described in these pages is to outline the benefit of binaural hearing in different kinds of applications, most of all linked with hearing aid and hearing loss problems. In such cases, the intelligibility of signals such as speech is highly influenced by the typology and the severity of the hearing loss, and also by the enhancement of binaural effects such as binaural masking

and the Cocktail Party Effect. However, hearing aid devices can greatly ameliorate the situation, and increase the intelligibility of the target signals.

3.6.3 Precedence effect

In a normal listening environment (not in free-field), the signal coming from a given source reaches the hearing system through an infinite number of different paths. First comes the signal covering the direct path, which obviously arrives first, to be followed by a series of reflections (called echoes), the delay and intensity of which depends on the typology of environment. Such factors include the distance of the walls from the listener, and the absorption coefficients of the walls. It may also happen that the sound energy of the reflections is higher than that of the direct signal. The human hearing system is nevertheless able to draw important distinctions between the direct and the reflected signals, thus delays do not seem detrimentally to interfere with the sound localization mechanisms.

Wallach (1949) and others investigated how the hearing system reacts to echoes. Their experiment consisted in sending through headphones two impulses to the ears, the first with a very small ITD, and the second with a larger one, but inverted. The results of this experiment can be summarised in the following points:

- The signals that reach the ears are interpreted as a single auditory event if the temporal distance between the two clicks is sufficiently small (for clicks, this distance needs to be in terms of 4-5 ms, while for complex tones up to 40 ms): this effect is called *echo suppression*.
- If the two signals are heard as a single auditory event, the localization of the stimulus is given mainly by the ITD of the first of the two clicks: this effect is known as *precedence effect*, or also as the *Haas effect* (Haas, 1951). It is therefore understood how the ability to localize the latter signal depends highly on its temporal distance from the former.
- If the distance between the two impulses is minimal (less than 1 ms), the precedence effect does not happen, and the sensation is therefore of a compromised localization. This effect is called *summing localization* (Blauert, 1996).
- If the second tone is sufficiently increased in terms of amplitude (10-15 dB), the precedence effect is cancelled, and two separate tones becomes clearly audible.

- The precedence effect can require some time in order to stabilise. If the two clicks are presented with a delay of 8 ms, it is clearly possible to differentiate them. If they are presented, for example, at a rate of four per second, after a few seconds it becomes impossible to continue to differentiate them. In the same way, fast changes in the acoustical conditions may cancel the precedence effect: inverting the ITD of the click, for example, changes therefore the sound image, and the precedence effect disappears for some seconds, only to re-appears after a certain number of repetitions.
- The precedence effect does not cause a complete loss of information given by the echoes. The listener can easily differentiate a sound with an echo from a completely "dry" one, and even small variations in terms of reflection can easily be perceived (when the hearing system analyses the reflections, it is in fact able to acquire acoustical information about the environment where the stimulus is presented).

The precedence effect is therefore very useful for determining the localization of a sound source in a reverberant environment. Nevertheless, it may occasionally show negative aspects. An example of this is when listening to single signal in a stereo loudspeaker system: delays longer than 1 ms between the two speakers can activate the precedence effect, with the result of a sound source being localized only in the loudspeaker from which the signal arrives first – and not in the other. This problem becomes more important when different subjects are listening to a more complex loudspeaker setup. Using an array with eight speakers, each placed at a corner of an octagon, with the listeners in the centre, it is clear that each individual will not be able to sit at exactly the same distance from all of the loudspeakers. A listener sitting, for example, two metres closer to the two loudspeakers on the left will perceive a stimulus coming from those loudspeakers 6 ms before hearing the same signal coming from the speakers placed on the right. Therefore, if the signal played back through the system is exactly the same over the eight loudspeakers, the perception of space will be significantly influenced by the position of the listener.

This effect is highly relevant when discussing sound reproduction systems, including binaural ones, more of which will be discussed in Chapter 9.

3.7 Brief summary

In the present Chapter, an introduction has been made to binaural phenomena regarding the perception of angles, starting with a description and analysis of differences in interaural time and intensity, and of direction dependent filtering. The mechanisms of the localization of sound sources placed in the three planes have been illustrated and discussed, analyzing the accuracy of these phenomena for both the azimuth and elevation angles.

In the concluding Section (3.6), a brief overview has been given of a selection of the most important binaural effects.

Chapter 4

4. Measurement of an HRIR Database

The mechanisms of spatial hearing have been described (Chapter 3), and basic notions about the simulation of a linear and time-invariant system have been discussed (Chapter 1). Attention can now move towards the simulation of three-dimensional soundscapes over headphones, which comprises the main topic of this research.

An introduction will be given to the measurement of the impulse response of a linear and time-invariant system. Next, the topic will proceed to the measurement of the IR from a dummy head system, reporting information on the measuring technique and system, azimuth and elevation sampling, and IR processing and editing.

Regarding the former part of the chapter, a paper on IR measuring techniques was presented at a conference in the first year of the Ph.D. (Picinali, 2006). The paper introduced no innovative techniques as it was merely a summary of techniques for IR measurement already introduced by other researchers; it set out to analyse the strengths and weaknesses of each in order to assist in the choice of the most suitable according to the typology of a task to be accomplished.

In contrast, the latter part of the chapter presents innovative experiments for the measurement of the HRIR database. When the experiments contributing to this thesis were carried out, the use of the sweep technique was exploited only for architectural acoustics tasks, while the best-known and most widely used HRIR databases were measured using other techniques, such as the MLS (Gardner, 1994; Algazi, 2001). Since then, only one HRIR database has been released using the sweep technique (the IRCAM Listen project[1]), but significant differences can be outlined between this example and the methodology described within this research thesis: the Listen database was measured with data provided by different individuals, not from a dummy head; furthermore, measurements were taken in an anechoic chamber, therefore no HRIR editing was required (*see* Section 4.4).

1 *See* http://recherche.ircam.fr/equipes/salles/listen

4.1 Measurement of the IR of a linear and time-invariant system

Chapter 1 provided notions about basic Digital Signal Processing principles (references will be given to the Basic Notions Chapter within this section). Here, a brief review is performed of different methods for the measurement of the impulse response of a linear and time-invariant system (Picinali, 2006).

4.1.1 The deconvolution

Considering a system *S*, knowing the input signal *x* and the signal *y* measured in output, *h* (the transfer function of the system) needs to be determined. In order to accomplish this it is necessary to find a signal *x* which has an inverse x^{-1} so that:

$$x \otimes x^{-1} = \delta$$

Hence, the following can be obtained:

$$y \otimes x^{-1} = h \otimes x \otimes x^{-1} = h \otimes \delta = h$$

This means that through knowing x^{-1} (starting from *x*) and the measured response *y*, it is possible to obtain *h*.

In theory, as much as in the digital domain, the measurement of the IR of a system is rather straightforward. Achieving it requires the input into the system of a *Diracδ* (a signal made by a one followed by a sequence of zeros), and the recording of the output signal. As stated in Chapter 1, the *Diracδ* has a flat spectral content. It is then possible to record the response of the system for all of the frequencies (in this case, taking into consideration the Nyquist theorem, the frequency range is limited to between 0 and samplerate/2). (Rabiner, 1975:296-300).

On the theoretical level, this operation seems to be simple; however, there are in practice numerous difficulties: firstly, it is essential to reproduce a sufficiently intense impulsive noise with at least 60 dB of Signal to Noise Ratio, and one that is short. For example, working at a sample rate of 96 kHz, the impulse should last for 1/96000th of a second.

An example could be given using a gunshot; unfortunately, this does not generate a signal with the duration of one single sample, but of a few tens of cycles. To circumvent this problem, the signal may be convolved with its inverse. This technique, known as

the *Time Reversal Mirror*, helps in arriving quite close to the *Diracδ*, although this is impossible to reach exactly due to the difficulties of calculating x^{-1} with sufficient precision.

Synthesizing a digital impulse then reproducing it through a loudspeaker could offer a more successful solution; however, with an elevated intensity for such a short time, the reproduction of the IR signal through a transducer, in this case a woofer and a tweeter, is almost impossible without frequency and phase distortions, and these, of course, would create relevant problems for a precise measurement of *h*.

4.1.2 The pink and the white noise

Considering the problems outlined in the previous lines, it could be convenient to pass from the time to the frequency domain using the Fourier transform. The convolution between two signals in the time domain becomes a simple multiplication in that of frequency:

$$y = x \otimes h \qquad H = X * H$$

Each frequency within the spectrum is multiplied with an *m* coefficient, for a result of *m* operations (in the time domain, they were m^2), considering then that the transforms and the anti-transforms have a limited computational cost (referring to the *Fast Fourier Transform*, which is much more efficient compared with the *Discrete Fourier Transform*, *see* Section 1.4). With the above technique, the measurement of the *m* coefficients is simple, because they represent the quotients between *X* and *Y*. Once the coefficients are obtained, it is sufficient to carry out an anti-transform to obtain *h(t)*. In this particular case, *H* is defined as the *transfer function*, while *h* is the *impulse response*.

Nevertheless, there is a problem of instability; where a frequency has a null *m* coefficient, its relating *H* coefficient diverges. To resolve this problem, it is possible to refer to the white and the pink noise signals. These are two kinds of signals that have the same energy on all frequencies. The white noise has a flat spectrum if displayed on a linear frequency scale, while the pink noise has a flat spectrum if displayed on a logarithmic frequency scale. The rich frequency content of these signals renders them particularly useful for many applications. Nevertheless, this solution also seems to be unsuitable; in

fact, the samples within these signals are generated randomly, therefore the spectrum is rather discontinuous if visualized with a short windowing and the phase is unknown. These frequency and phase problems make this technique unsuitable for the majority of cases, where the resolution of the phase and the frequency is essential.

4.1.3 The MLS signal

A particularly "clever" signal that could be used instead of the white and pink noises is the MLS (*Maximum Length Sequence*, *see* Figure 1 and 2) (Rife, 1989). It is a binary sequence generated by a shift register that follows the scheme shown in Figure 1, below.

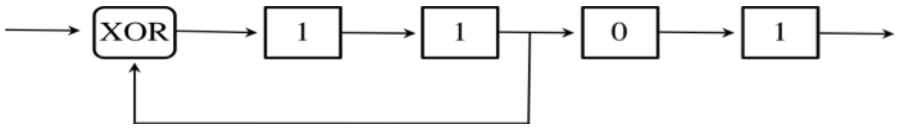

Figure 1. The MLS generation scheme

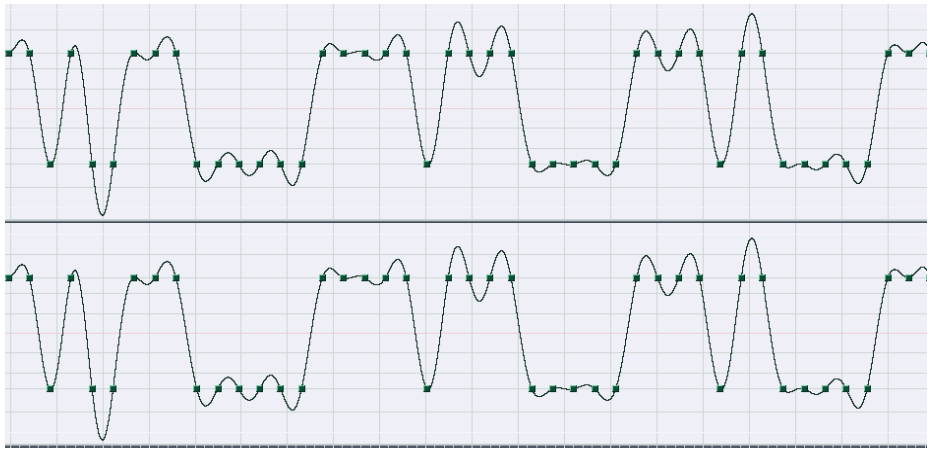

Figure 2. An example of a stereophonic MLS signal generated using the Aurora Plugins[2]

With a correct positioning of the XOR, different kinds of MLS signal may be obtained. A highly important property of this signal is that, by correlating it with itself (auto-correlation), (Rabiner, 1975: 401), it is possible to obtain a *Diracδ* without using the FFT algorithm. In fact, generating an MLS *x* signal to be input into the system would be

2 *See* http://www.aurora-plugins.com

sufficient, sampling the *y* output signal and cross-correlating *x* with *y*; this operation, in the time domain, will generate the impulse response *h*.

$$If \quad y = h \otimes x \quad and \quad x \subset\subset x = \delta \qquad ==> \qquad y \subset\subset x = h \otimes x \subset\subset x = h \otimes \delta = h$$

Unfortunately, the principal disadvantage of this technique is its strong dependence on the linearity of the system to be measured. The MLS technique requires a perfectly linear and time-invariant system (Svensson, 1999; Paulo, 2008). Inexistent echoes and phase problems can appear even with minor non-linearities. Such problems make this simple technique unusable for the IR measurement where an utterly precise measurement system is impossible to achieve; in the analogue domain, this happens frequently.

4.1.4 The sinus-logarithmic sweep signal

It seems that the most effective and efficient current technique for IR measurement is that which uses a signal made by a sinusoidal function moving from the low frequencies to the high; this generates a pure tone that increases its frequency with time. Swept frequency sinusoids have a long history in room measurement. It is difficult clearly to determine who first used them. Nevertheless, it is possible identify a list of relevant publications (Berkhout, 1980; Muller, 2001; Griesinger, 1996, and Farina, 2000).

The advantage of this technique is that generating the sweep signal *x*, its inverse x^{-1} is simply the *x* signal reversed on the time axes. Knowing x^{-1} and measuring *y*, it is possible to calculate the IR *h* with a simple deconvolution operation:

$$y \otimes x^{-1} = h \otimes x \otimes x^{-1} = h \otimes \delta = h$$

The only drawback of this calculation is that the convolution operation is not streamlined, and the computational efficiency of the sweep technique is lower, as compared with those described above. The problem is, however, not particularly relevant in this case, due to the fact that these convolution operations can be batched and run offline just after the measuring session.

The sweep signal may be linear or logarithmic, depending on the frequency-increasing curve. The most frequently used is the logarithmic because with it, greater energy may be invested in the lower frequencies (critical zone), and the higher ones run faster.

The sinus-logarithmic sweep formula is the following:

$$x(t) = \sin\left[\frac{2\pi f_{\inf} \cdot T}{\ln(\frac{2\pi f_{\sup}}{2\pi f_{\inf}})} \cdot \left(e^{\frac{t}{T} \cdot \ln\left(\frac{2xf_{\sup}}{2xf_{\inf}}\right)} - 1\right)\right]$$

where f_{inf} is the starting frequency, f_{sup} is the arrival frequency and T is the time duration.

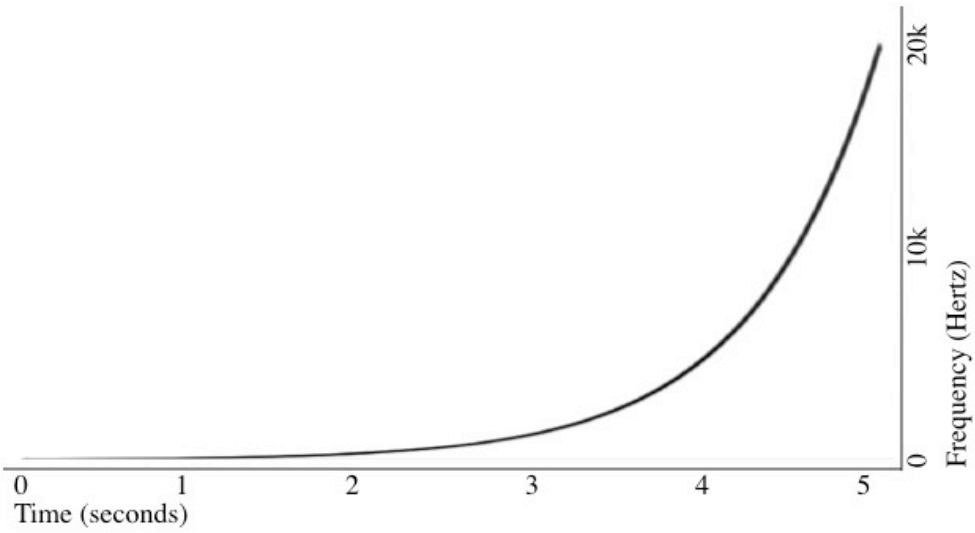

Figure 3. The sonogram of a sinus-logarithmic sweep signal

4.2 The dummy head and the measuring system

Considering the dummy head (an artificial head mannequin with two microphones placed at the entrance of the ear canals) to be a linear and time-invariant system to be simulated, it is possible to measure the HRTF measuring all the HRIRs for all the possible positions of a sound source around the listener. This can be done using the techniques described in the previous section (Section 4.1.4), in particular the sinus-logarithmic sweep.

In the next sections, the dummy head system will be presented; information on the measurement of the HRIR will be given in terms of both the choice of measuring system, and of the sampling of the angles of azimuth and elevation.

4.2.1 The Dummy Head

Various dummy heads are available on the market.[3] For this thesis, the decision was taken to use a custom-built dummy head for the measurement of the HRTF. Irrespective of the cost of a new dummy head system (>£10k), using a custom-built dummy head allowed greater flexibility in the measuring system; it provided the opportunity of choosing the microphones and their placement within the ear canal, and of changing the pinna mould quite easily for possible future experiments into HRTF with individual characteristics.

Figure 4 shows the dummy head built for this specific project. The head itself is made of polystyrene covered with a 4 mm layer of latex. The pinna moulds were made using a particular bi-component material in use by audiometrists (Dreve Otoform Ak[4]) for the negative, and latex for the positive and final product (Figure 5 shows both the negative and the positive moulds). For this specific experiment, the pinnas used in preparing the mould were those of the author; in future experiments, it will be relatively quick (about 48 hours) to prepare moulds (positive and negative) from different pinna shapes.

Another important factor in the construction of the dummy head is the positioning of the microphones (about which further data will be given later in this section). As Hammershoi (1995) suggested, it was finally demonstrated that all spatial information (the localization cues) regarding the location of the sound source around the listener are already present in the signal at the entrance to the ear canal. Chapter 3 of this thesis has outlined how the resonances of the ear canal are independent of the angle of incidence of the signal, whereas all of the resonances of the pinna are, further, highly dependent on it. It has therefore been considered appropriate to place the microphones inside the dummy head exactly at the beginning of each of the ear canals (*see* Figure 6).

Another important factor considered is the simulation of the shoulders and torso of the dummy (for the importance of these in the spatial hearing process, *see* Chapter 3). A sweater filled with packaging material was placed exactly under the dummy head, in order properly to simulate the presence of the shoulders and torso (*see* Figures 20 and 21).

3 *See*, for example, the Neumann KU 100 or the Kemar Dummy Head Microphone

4 *See* http://www.dreve.de

Figure 4. The custom-built dummy head

Figure 5. The positive and negative pinna moulds

Figure 6. The placement of the microphones at the entrance to the ear canal

4.2.2 Sampling of the azimuth and elevation angles

In order to measure the HRTF of a dummy head, HRIRs need to be measured from different positions of the sound source; thus, the azimuth and elevation parameters need to be sampled around the head. For each of the sampled positions, then, the IR needs to be measured. The sweep signal must therefore be reproduced through a loudspeaker and recorded from the two microphones placed inside the dummy head: then, in the following stage, all recorded sweeps will be convolved with the inverse of the original sweep, giving as a result the HRIRs corresponding to the positions of those specific loudspeakers.

In the literature, different azimuth and elevation scales used for various HRTF measurement experiments may be found. As a reference, an overview can be made of the research projects of MIT Kemar (Gardner, 1994) and CIPIC (Algazi, 2001):

- MIT: the space around the dummy head is sampled at elevation angles from –40° (40° below the horizontal plane) to +90° (directly overhead). At each elevation, the full azimuth range (360°) is sampled in equal-sized increments. The increment sizes were chosen in order to maintain homogeneity in the distribution of the HRIR on the

surface of the sphere around the dummy head. A total of 710 locations were sampled and are given in Table 1:

Elevation	Number of Measurements	Increment of Azimuth
-40	56	6.43
-30	60	6.00
-20	72	5.00
-10	72	5.00
0	72	5.00
10	72	5.00
20	72	5.00
30	60	6.00
40	56	6.43
50	45	8.00
60	36	10.00
70	24	15.00
80	12	30.00
90	1	x

Table 1. MIT Kemar HRTF measurement experiments, number of measurements and increments of azimuth at different elevations (After Gardner 1994:3)

- CIPIC: the sampled positions are specified in interaural-polar coordinates (the interaural-polar azimuth is limited to the range from –90° to +90° and points that are behind the subject are found at 180° elevation). Elevation is uniformly sampled in 360/64 = 5.625° steps from –45° to +230.625°; in order to obtain roughly uniform density on the sphere, azimuths were sampled at –80°, -65°, -55°, from –45° to 45° in steps of 5°, at 55°, 65° and 80°. This leads to spatial sampling on 1250 points, as illustrated in Figure 7.

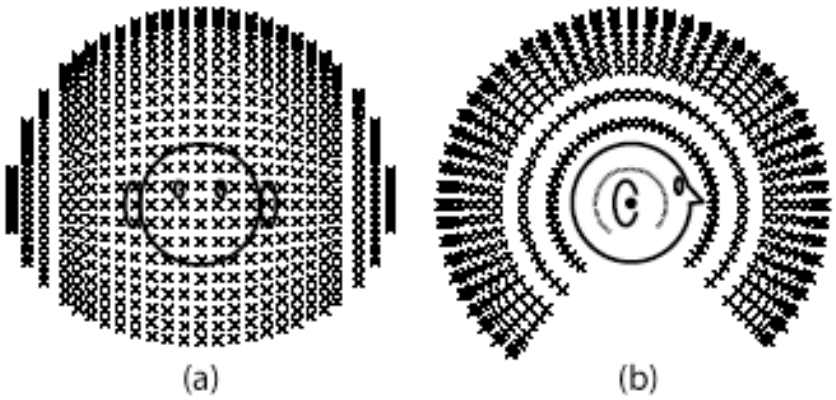

Figure 7. CIPIC HRTF measurement experiments, location of data points (a) front (b) side (After Algazi, 2001:1)

In order appropriately to select the azimuth and elevation sampling scale, the MAA or Minimum Audible Angle (*see* Chapter 3) for all three planes needs to be considered accurately.

It may be noticed that the MAA varies in function of the angle where is measured, and in function of the frequency of the signal used for the test: at 0° it goes from one to three degrees, while at 75° from seven to fifteen degrees. Regarding the frequencies, it is necessary to make approximations, while regarding the respective sensitivity at the various angles, an *ad hoc* sampling scale may be elaborated.

After several analysis and considerations, a possible sampling scale for the azimuth parameter is proposed:

- From 0° to 10°, in steps of 2° (with a single step at 5° to maintain the compatibility with an azimuth scale with fixed steps of 5°).
- From 12.5° to 60°, and in steps of 2.5°.
- From 65° to 180°, in steps of 5°.
- The same can be repeated specularly from 180° to 360°.

According to Algazi (2001), in order *"*to obtain roughly uniform density on the sphere" around the head it is necessary to use a different azimuth sample scale for each of the different elevations. Obviously, at 80° of elevation it will not be necessary to sample the azimuth each 2.5°. On the other hand, it would be overly complex to employ a different azimuth scale for each of the elevation measurements.

In the opinion of the author of this thesis, in order to maintain the uniform density of the sample positions on the sphere, but at the same time to keep the whole process simple, the optimum choice lies in using two different azimuth scales: that outlined here for elevations between –40° and +40° (extremities included), and a fixed step scale (of each 5°) for elevation at –60°, -50° and from +50° to +80°.

All calculations described previously would generate a complex scaling system, with different steps for both the azimuth and the elevation angles, creating therefore problems of compatibility between this and other HRTF databases; it is also certain to make the measurement process more complex. Furthermore, the MTIRC (Music, Technology and Innovation Research Centre, where this Ph.D. research has been carried out) agreed to buy, specifically for these experiments, an automatic turntable, the Outline ST2 with the ET2 electronic controller. It was then used for automating the HRIR measurement process: the turntable can move automatically in steps of 2.5°, a minimum movement incompatible with the sampling scale described above, while it can be synchronized with a PC for a perfect automation process (*see* Figure 6). When the dummy head and its stand are placed on the turntable, this allows its rotation at given steps and the loudspeaker to be left fixed in one position. The height and orientation of the loudspeaker were then changed for the measurement of the HRIRs at different elevation angles (*see* Figure 18).

Finally, this choice (of fixed 2.5° steps) seemed to be the more acceptable in terms of both measurement precision and complexity. Table 2 shows a recapitulation of the azimuth and elevation scales; a total of 1729 locations have been sampled.

Figure 8. The Outline ST2 turntable, with the ET2 controller

ELEVATION	AZIMUTH	SAMPLED POSITIONS
-50° and –60°	Fixed steps, each of 5°	144
From -40° to 40°	Fixed steps, each of 2.5°	1296
50°, 60°, 70° and 80°	Fixed steps, each of 5°	288
90°	One sample	1

Table 2. Recapitulation of the HRIR measurement positions and azimuth increment at each degree of elevation

4.2.3 Other choices of parameter

Here follows a list of other parameters and values chosen for the HRIR measurement experiments:

- Distance between the dummy head and the loudspeaker: 1 metre. This is standard for HRIR measurement experiments (Algazi, 2001). Nevertheless, experiments for the HRIR measurement at different distances have been carried out in following stages of this research (*see* Chapters 5 and 6).
- Height between the dummy head and the floor: 165 cm, calculated at the two ear levels. This represents the average height of a human being.

- Sweep length and frequencies: the choice of a 10-second sinusoidal sweep signal from 22 Hz to 22 kHz was made, given different standard values in architectural and environmental acoustic (*see* Capra, 2002, and Farina, 2005).
- SPL calibration level: the loudspeaker and amplification system is calibrated in order to generate an SLP of 94 dB at 1 kHz at the position of the head, and the microphones and preamplifiers are calibrated in order to generate a level of -6 dBfs (measured in the DAW application) for that specific SPL. These values have been chosen according to standards in acoustic measurement experiments.

4.2.4 The measuring system

Here follows a list of the equipment used for the measurement of the HRIR.

- Audio interface: Motu Traveller Firewire[5] (working at 96 kHz and 24 bit, with a unique clock synch for both reproduction and recording).
- Microphones: two DPA 4060bm[6], omni-directional condenser microphones (*see* Figure 9 for the frequency response).
- Phonometer: Phonic PAA2[7].
- Loudspeaker: GENELEC 8040A[8], two-way biamplified loudspeaker (*see* Figures 10 and 11 for technical data).
- Software:
  - Apple (ex E-Magic) Logic Audio Pro 7.1[9].
  - Adobe Audition 1.5[10] (with Aurora Plugins[11], for the sweep generation and for the convolution process)

5 *See* http://wwwmotu.com

6 *See* http://www.dpamicrophones.com

7 *See* http://www.phonic.com

8 *See* http://www.genelec.com

9 *See* http://www.apple.com/logicstudio/

10 *See* http://www.adobe.com/products/audition/

11 *See* http;//www.aurora-plugins.com

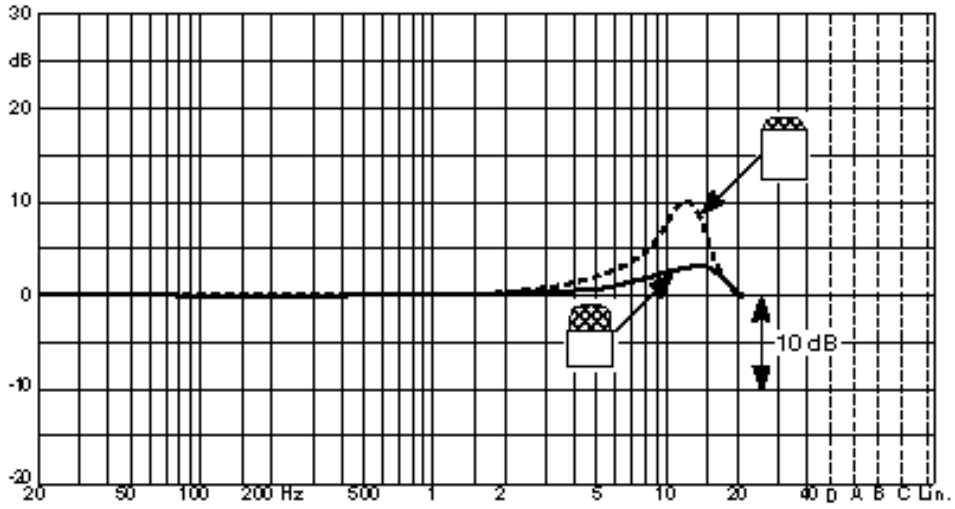

Figure 9. Frequency response diagram of the DPA 4060bm microphone. For this experiment, the lower cap has been used (the reference curve is the lower one, with a flatter response above 5 kHz). (Data taken from www.dpamicrophones.com)

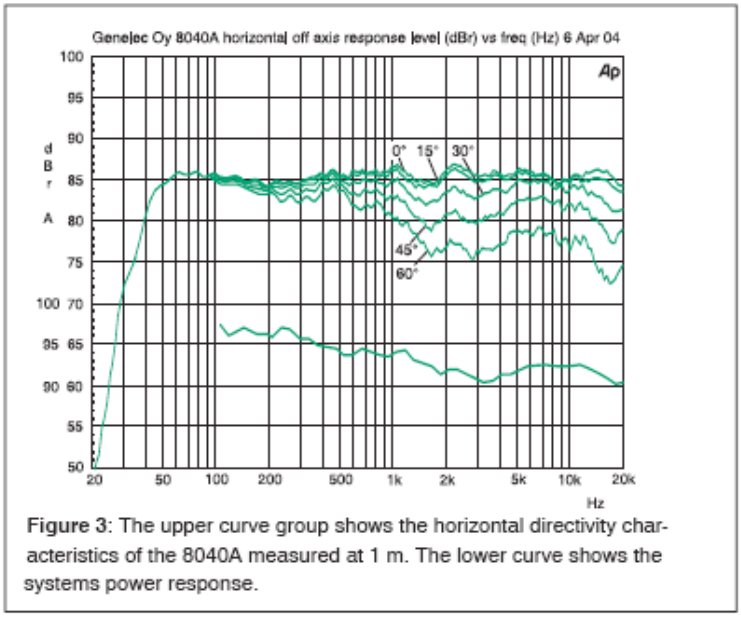

Figure 3: The upper curve group shows the horizontal directivity characteristics of the 8040A measured at 1 m. The lower curve shows the systems power response.

Figure 10. Frequency response diagram of the Genelec 8040A loudspeaker. (Data taken from www.genelec.com)

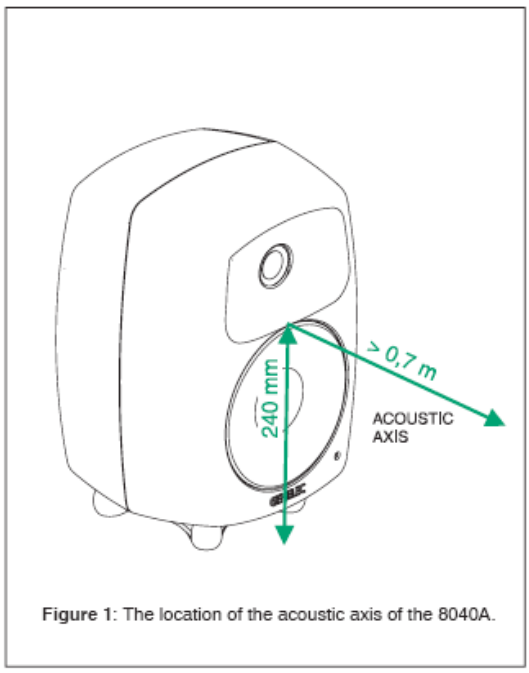

Figure 1: The location of the acoustic axis of the 8040A.

Figure 11. The location of the acoustic axis of the Genelec 8040A loudspeaker. This axis has been used in order properly to align the loudspeaker with the ears of the dummy head. (Data taken from www.genelec.com)

4.3 Calibrations

This section discusses the choice of the room for the experiments, and the calibration of the measuring system used in the experiments.

4.3.1 The room

In the literature, experiments for the measurement of the HRTF have always been carried out in anechoic environments (for example, Algazi, 2001, and Gardner, 1994). Although an anechoic chamber was not available in the facilities of the DMU, an environment called the “Diffusion Room” was made available for the performance of the experiments.

This is a 10 x 5 x 2.5 m room with absorbent materials on walls, ceiling and floor. It has an RT60 of 0.3 seconds (+- 0.1 sec across the whole frequency range), insufficiently short to be considered anechoic. Nevertheless, given the dimensions of the room, it may be estimated that by placing the dummy head in the centre of the room, the first reflections would not arrive earlier than 3-4 ms from the direct signal, and these could

therefore easily be removed from the final impulse file (further discussion of this topic will be given in Section 4.4).
Different rooms have also been used in determining the HRTF measurements performed within this research, as will be shown in Chapters 5 and 6.

4.3.2 About calibrations

The measurement experiments will carry a series of alterations due to the non-flatness, in terms of frequency response, of each measuring system used. For the transmission of the measurement signal (a sinus-logarithmic sweep) a loudspeaker will be used with a non-flat frequency response, in a non-anechoic chamber (again, with a non-flat frequency response); the signal will be recorded with two microphone capsules, followed by two microphone preamplifiers and two A/D converters. These are not, of course, flat in the frequency response.
In summary, the signal to be recorded will be the sum of the transfer functions of all of these systems, necessarily including the transfer function of the dummy head, or the HRTF, the ultimate goal of the measurement. Rendering this as a mathematic formula, it can be written:

$$I(t) \otimes A(t) \otimes H(t) \otimes M(t) \otimes P(t) = F(t)$$

This is where *I(t)* is the digital synthesized impulse, with a known frequency response; *A(t)* is the transfer function of the loudspeaker and of the room; *H(t)* is the HRTF; *M(t)* is the microphone capsule transfer function; *P(t)* is the transfer function of the microphone preamplifier and of the A/D converter, given that the microphone preamplifier is absolutely necessary because the impedance of the microphones and their output voltage cannot be directly input into the A/D converter, and *F(t)* is what will be recorded by the microphones. The result is the convolution of the original impulse with the IR of all the systems cited thus far. The element that has to be measured is *H(t)*. *I(t)* is known, as is *F(t)*, and therefore it is necessary to find *A(t)*, *M(t)* and *P(t)*. The alterations brought about by the measuring system in terms of frequency response need to be established in order to achieve the accurate measurement of the HRTF only.
Two different approaches may be used in this case: the first is to measure the frequency response of each of the single parts of the system ("sub-systems"), and to make an

inverse filter for the inversion of each of these; the second is to consider all factors together, and to make an inverse filter for the whole system.

The choice between these two methods depends on the kind of experiment to be conducted. Here, it is not essential to know the frequency response of every single element of the measuring system. In mathematical terms, it is not essential to know *A(t)*, *M(t)* and *P(t)* individually, whereas in order to measure *H(t)* it will be sufficient to know *C(t)*, the convolution among these three transfer functions:

$$C(t) = A(t) \otimes M(t) \otimes P(t)$$

In order to perform this calibration, the response of the measuring system only needs to be measured. The two microphones were therefore placed on a stand in the exact position where the dummy head would be located during the experiments, the loudspeaker was placed in the 0° az and 0° el position, and an IR was measured using the sweep technique (*see* Figures 12 and 13).

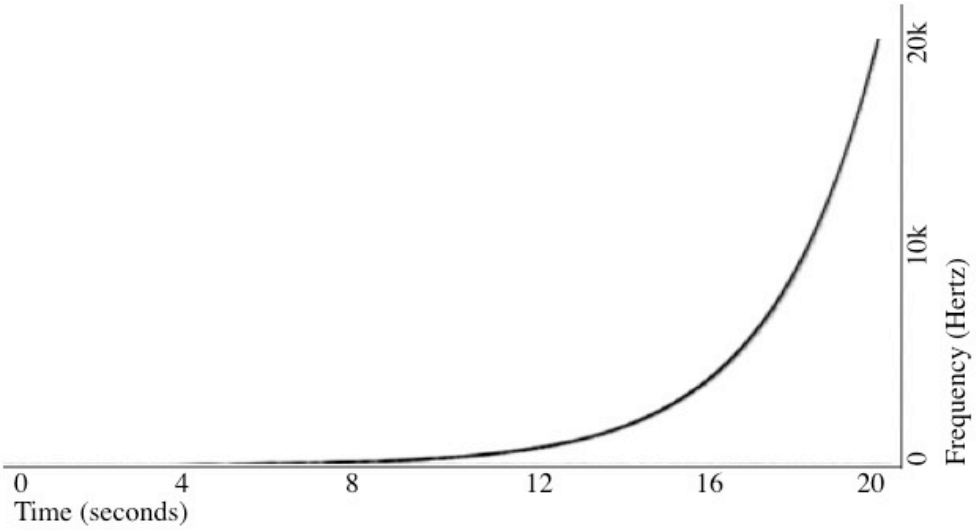

Figure 12. The sonogram analysis of the averaged (left and right channel) sweep recorded from the measuring system

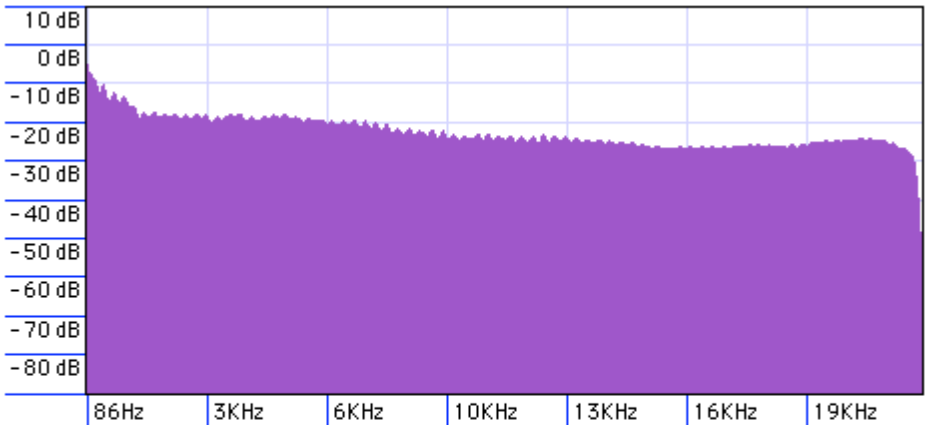

Figure 13. The spectrum of the averaged sweep (left and right channel) recorded from the measuring system

The curve in Figure 13 demonstrates that there are differences within +-7 dB within the whole spectrum, leaving aside the small *ripples* (alterations can also be generated by the Fourier transform parameters used for passing from the time to the frequency domain). Even if this non-flatness in the frequency response seems not to be particularly relevant, two attempts have been made to create an inverse filter. The first attempt employed the MatchEQ plugin (from Logic Audio 7.1) in Figure 14:

Figure 14. Inverse equalization curve generated by the MatchEQ plugin within Logic Audio 7.1.

As shown in Figure 14, the MatchEQ processor, input with the measured sweep signal, generated an inverse filter that may be applied on the original sweep signal in order to compensate for the non-flatness of the measuring system. When this filter is applied on the original sweep signal, the resultant frequency response of the whole measuring system seems clearly to be flatter when compared to the frequency response shown in Figure 13, as is shown here in Figure 15:

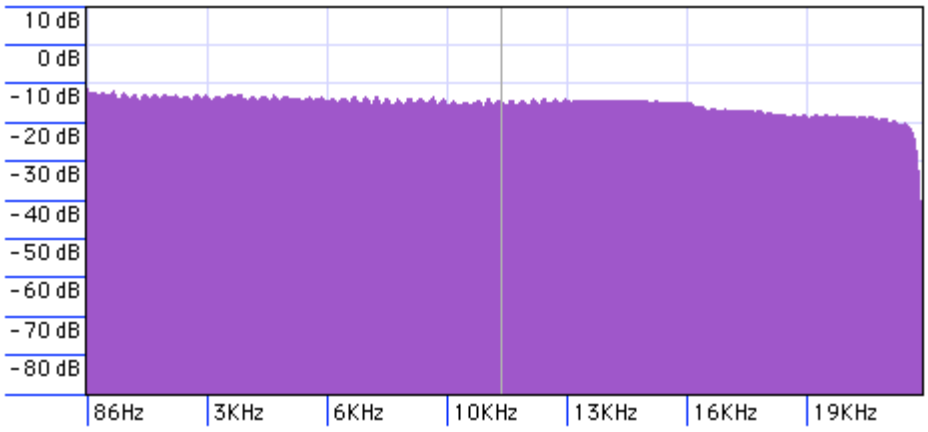

Figure 15. The spectrum of the averaged sweep (left and right channel) recorded from the measuring system and applying the inverse filter generated by the MatchEQ processor.

Nevertheless, after analysing the sonogram of the measured sweep signal processed with the inverse filter, unexpected and problematic results in the spectrum towards the end of the sweep signal, where the inverse filter applies a +13 dB boost, were noticed, as Figure 16 demonstrates:

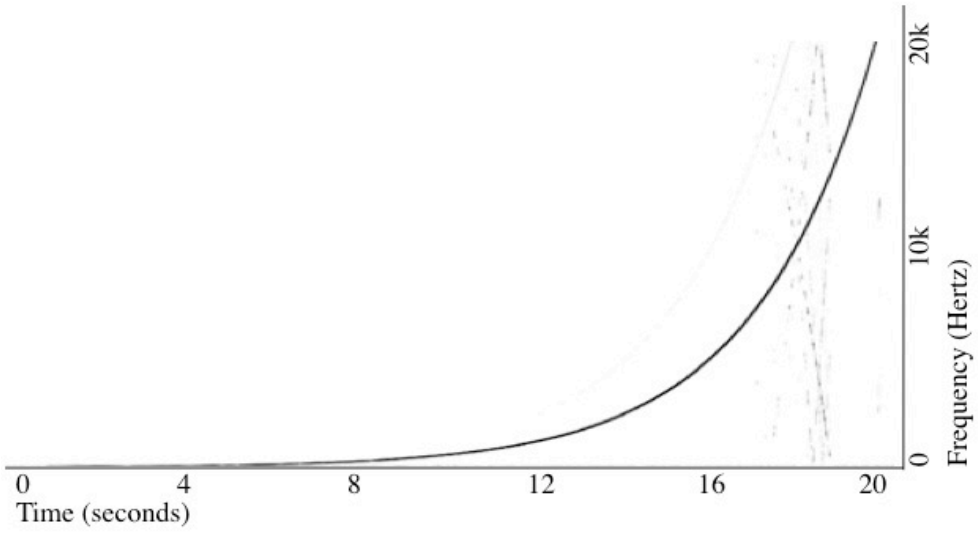

Figure 16. The sonogram analysis of the averaged (left and right channel) sweep recorded from the measuring system and applying the inverse filter generated by the MatchEQ processor.

Given the fact that such results were absent from the sonogram analysis performed on the unprocessed recorded sweep signal (*see* Figure 12), they may be considered as errors generated by the MatchEQ processor.

The second attempt was made by generating an inverse impulse response from the actual IR corresponding with the response of the measuring system. The inverse IR was then convolved with the measured HRIR, in order to cancel the frequency and phase distortions introduced by the measuring system and, therefore, to isolate the elements of the HRTF.

The inverse filter was created using the Inverse Filter Aurora plugin in Adobe Audition 1.5. Figures 17 and 18 show the stereo waveforms of the original IR and of its inverse. To verify the accuracy of the filter inversion operation, the two IRs have been convolved; the waveform of the resultant IR can be found in Figure 19. The resultant IR should be a simple *Diracδ* in order for the process to be considered free from any kind of frequency coloration. A series of zeroes, a one, and then zeroes again should result. Leaving aside the ripples generated by the interpolation calculated from the audio editor, Figure 19 shows how the inversion is almost perfect for the left channel, yet with a few more problems for the right channel. However, informal listening tests carried out

on the HRIR both before and after inverse filtering determined that such negligible deviations have little relevance in terms of the accuracy and realism of spatialization.

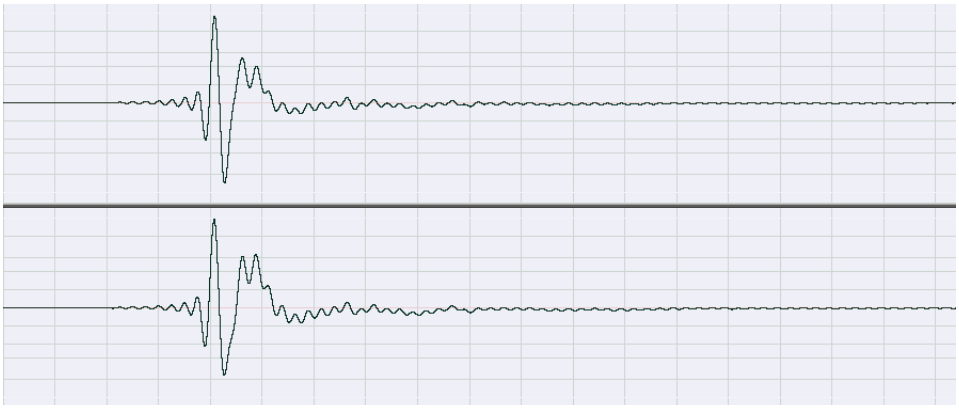

Figure 17. The waveform of the measuring system IR

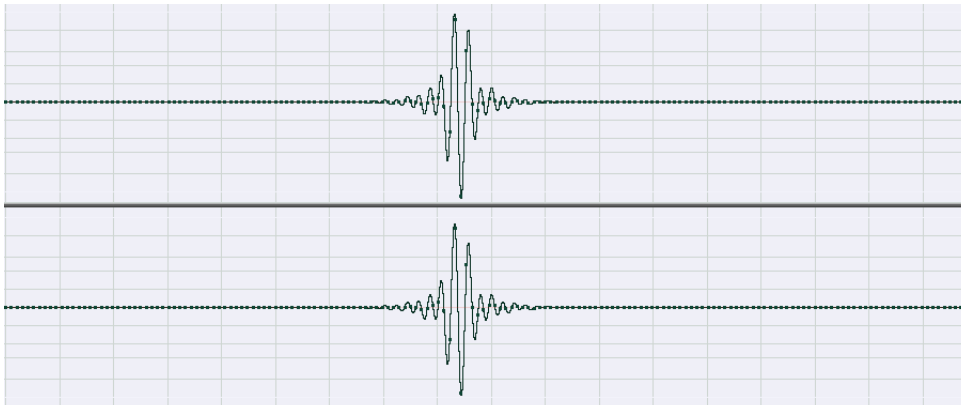

Figure 18. The waveform of the measuring system inverse IR

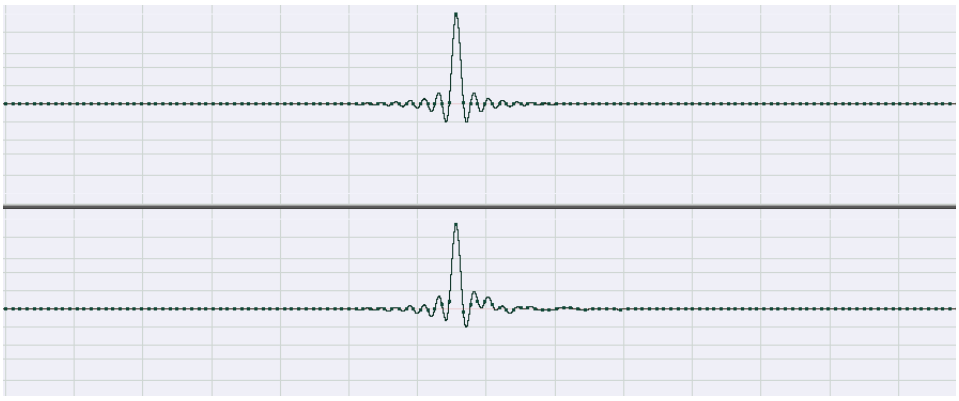

Figure 19. The waveform of the measuring system IR after the convolution with its inverse

Every inverse filtering operation carried out in order to flatten the frequency response of the measuring system will anyway create frequency and phase distortions; an example is the MatchEQ process. Even if most of these were excluded from the IR thanks to the use of the sinus-logarithmic sweep (*see* Section 4.1.4), it has nevertheless been considered appropriate to apply inverse filtering neither before nor during the HRTF measurement process. However, the response of the measuring system (the microphones without the dummy head) has been measured and stored, as have the results using the inverse filter, in order to allow an eventual post-processing stage for the "inversion" of the total of the measured HRTFs. A perceptual test to determine whether the inversion process results in a better performance in terms of sound spatialization may also follow.

A brief discussion should now be offered. The frequency alterations introduced by the measuring system are not direction-dependent, and therefore do not vary according to the position of the sound source around the dummy head, in contrast to the DDF. Considering the plasticity of the human hearing system (Parks, 2004), and its ability in adapting relatively quickly to new situations, most of these direction-independent alterations related to the frequency and phase response of the measuring system may be considered irrelevant from the perspective of spatial hearing. After a few minutes of listening to the binaural simulation performed using the measured HRTF, the hearing system should be able to adapt to the new situation (in this case, to the new frequency response) and automatically to ignore the alterations introduced by the measuring system. Further elaborations on this have already been carried out in the introductory chapter.

4.4 HRIRs measurement and editing

This section will address the measurement sessions, as well as the calculation and editing of the IR.

4.4.1 The measurement sessions

Figures 20 and 21 show selected photographs from the HRTF measurement experiment. Apart from some minor issues of calibration and distortion (all of which were resolved before the actual measurements), it may be claimed that the four experiment sessions, carried out on four different days, were utterly satisfactory.

Figure 20. A photograph from the HRTF measurement experiments, with the loudspeaker placed at 0° Az and 0° El.

Figure 21. A photograph from the HRTF measurement experiments, with the loudspeaker placed at 90° Az and 20° El.

4.4.2 The IR editing

The recorded sweep signals were then convolved (using a batch processing script on Adobe Audition) with the inverse of the original (dry) sweep, generating the HRIR corresponding to the specific position of the loudspeaker and orientation of the dummy head during that measurement.

Figure 22 shows the oscillogram of the left and right channels from the HRIR for 0° Az and 0° El. Further to the topic discussed in Section 4.3.1, it may be noticed how the direct signal is perfectly distinguishable from the reflected one: at second 0.005 it is clearly possible to see the arrival of the first reflections coming from the floor and, subsequently, the arrival of the reflections coming from the walls. For this reason, the IR have all been edited in order to achieve two different HRIR databases. The first, unprocessed, contains the effects of the room (with a short reverb), and the second is pseudo-anechoic, for which the tails of the impulses have been cut at 0.004 seconds and the IRs have been faded using a cosine curve.

A normalization operation was also carried out on all of the HRIRs, resulting in a boost of +6 dB, thus a multiplication by factor 2 of all of the samples of the IRs, a value coherent with the calibration values described in Section 4.2.3. This calculation has of course maintained unvaried the relative gains between the Left and Right channels, and between the different HRIRs; therefore, the +6 dB boost was applied to all of the HRIRs independently of their amplitude, which was below -6 dBfs.

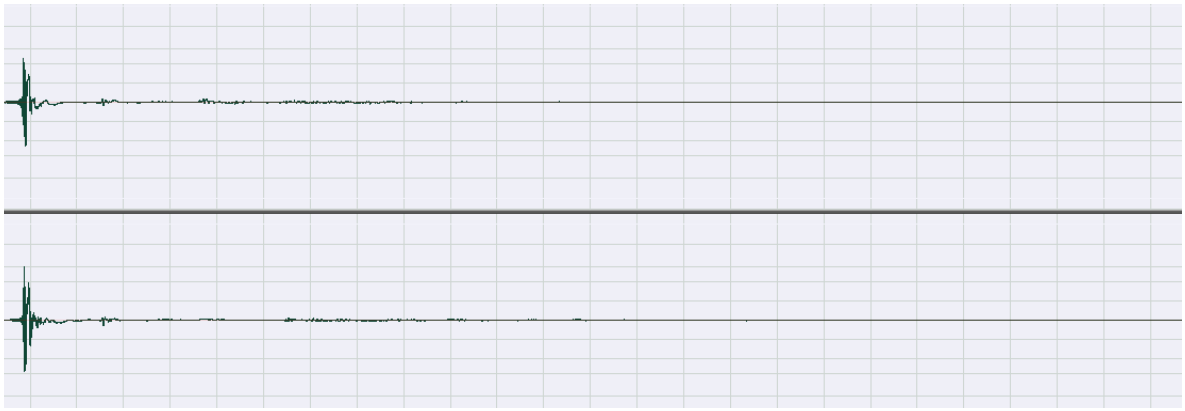

Figure 22. The waverform of the left and right channels from the HRIR for 0° Az and 0° El.

The HRIR database has therefore been created and organized using the following naming standard: *1m_0El_0AzANEC.aif*, where *1m* stands for the distance at which the HRTF has been measured; *0El and 0Az* for the angles of elevation and azimuth in degrees, and *ANEC* for the fact that that specific impulse has been edited in order to become *pseudo-anechoic*. Also, a second database has been created with the non-processed HRIRs, without the description *ANEC* in its title.

4.5 Brief summary

The present chapter has introduced the techniques for the measurement of the IR from a linear and time-invariant system. The advantages and disadvantages of each of the techniques described have been outlined, leading to the choice of the optimum technique for the measurement of the IR from a dummy head system (HRIR).

In the second part of the chapter, a description was given of the experiments carried out within the Ph.D. research for the measurement of an HRIR database, reporting carefully the characteristics of the measurement system, the different calibrations and editing performed during and after the experiments, and the final organization of the HRIR database.

Chapter 5

5. Binaural Phenomena for the Perception of Distance

Following the study into the binaural phenomena relevant to the perception of the angle of incidence of a given sound (Chapters 3 and 4), the topic will now move to the perception of distance, i.e., how the human hearing system is able to determine the distance of a given sound source from the listener. As will be outlined in the following sections, the perception of distance is an extremely complex process involving numerous different parameters of the sound input into the hearing system. Many of these parameters, such as the intensity of the sound source and the reverberation generated by the reflections of the sound on the environment, can also vary independently of the actual distance between the source and the listener, rendering their estimation especially complicated.

Before beginning the description and analysis of these phenomena, two factors need to be considered. The first is linked with the sluggishness of the human hearing system regarding the estimation of distance. When compared with the perception of the angle from which a sound is arriving (*see* the description of the Minimal Audible Angle or MAA in Section 3.5.4), the hearing system is particularly weak in terms of linear error of the estimation of distance (further information on this aspect will be given in Section 5.1.1). The precision of perception becomes even weaker for sources at more than one metre from the listener, to the extent that it may be almost impossible to establish with any accuracy the distance from an unfamiliar sound source (of which the apparent volume is not known) for distances greater than six or seven metres. While it is true that the factors influencing the perception of the angle, such as the familiarity of the listener with the signal and the acoustic characteristics of the environment where the signal is presented, are indeed important, their influence is far stronger as regards the perception of distance and make this, as stated, a considerably more complex yet, paradoxically, less precise process. Parameters, such as reverberation and the directional perception of the early and late reflections, other than the distance and the direction of a sound source are related to the spatial hearing process. Such parameters often apply the same cues used for sound source localization, yet they are more difficult to measure and, most

significantly, are rarely quoted as measured in the literature. Further reference to this will be made in Chapter 10.

The second factor that must be taken into account is the fact that the literature on the perception of distance is far less consistent than that on the perception of the angle. Much research upon which current knowledge is based was carried out fifty to one hundred years ago; it has never been repeated. Further experiments and tests are required in order better to understand the mechanisms performed by the human hearing system for the estimation of distance (for possible future experiments, *see* Chapter 10).

It should be noted that within this and the subsequent chapter (Chapter 6) the main innovations of this research work are described and analysed. The present chapter consists mainly in a review of the literature on the perception of distance (except for Section 5.3, where an actual original study is described); Chapter 6 will focus on the simulation of distance cues and on the creation of a binaural reverb algorithm, both of which may be recognised as the main innovations presented within this research.

5.1 Binaural perception of distance

In this section, the binaural perception of distance will be described, considering a broadband noise sound source placed at different distances from the head and analysing how the hearing system is able to determine the distance of the sound source. A description of the various pieces of research carried out on the IHL effect (the Inside the Head Locatedness effect, introduced in Chapter 1) will also be given in the attempt to outline when and why this effect appears. Further information about this topic can be found in Blauert (1996:116-137).

5.1.1 The perception of distance

The distance of the auditory event is calculated from the median point of the axis that passes between the ears. When a sound source is localized inside the head (IHL), the distance of the auditory event is smaller than the radius of the head. This is the case, for example, when listening to standard signals over headphones. It is important to underline that the body of knowledge of the mechanisms for distance hearing perception (important experiments on the performances for the estimation of distance can be found in Coleman (1962; 1963)) is considerably less complete than the knowledge of directional hearing topics (Blauert, 1996:117).

Presenting a subject with a broadband signal, with the sound source placed in the median plane, the distance of the auditory event coincides more or less with the one of the stimuli. However, this is not always the case with narrow band signals (Blauert, 1996:37-50). An increase in the distance of the sound source will correspond to an increase in the distance that is perceived, and vice versa. There are therefore attributes of the signals at the entrance of the ear canal that strictly depend on the distance of the sound source. The hearing system is able to decode those attributes and achieve the sensation of distance. As will be shown in the following lines, an absolute perception of distance does not exist. The hearing system works on a comparison among both other stimuli that arrive at the ear and of which the distance is known, and previously memorized sound material.

After the work of Blauert (1996:118), a classification scheme can be offered, reflecting those attributes of the ear input signals that depend upon the distance from the sound source:

1. At an intermediate distance, of between three and 15 metres, the sound pressure at the eardrum depends on the distance following the 1/r law. The sound pressure level (SPL) falls by one half (that is, a reduction of approximately 6 dB) for every doubled distance. However, this is valid only in free-field conditions, as will be seen in Section 5.2. This seems the only attribute on which the hearing system can establish the distance of the sound source.

2. At a greater distance, i.e., of more than 15 metres, the path that the sound wave needs to cover in the air cannot be considered distortion-free. Together with the decrease of sound pressure commensurate with an increase in the distance, an additional frequency-dependent attenuation is present, due to the fact that higher frequencies are attenuated more than lower ones. Therefore, spectral variations, as well as amplitude variations, are used for the estimation of the distance for distant sound sources.

3. Close to the sound source, at less than three metres, the effect of the curvature of the wavefront needs to be considered; the source must be assumed to be a point radiator, with a circular waveform that at greater distances may be approximated as being plane. The linear distortion caused by the presence of the head and of the pinna va-

ries according to the distance of the source, generating spectral variations differing from those generated for greater distances.

4. For the signals presented over headphones, the effect of the head and of the pinna is bypassed, and the perception is, frequently, unless spectral cues are deliberately added, of a sound source located inside the head (IHL).

Reverting to the first point made, a sound source far enough to consider the wavefront as being plane yet close enough for approximating the air absorption as being frequency independent, the only cue available for estimating the distance of the source is the scaling of the sound pressure level at the eardrum. Obviously, in order for the hearing system to establish the distance of the source with only the loudness cue, the sound pressure needs to be known and constant in time, that is, known in the context that the listener needs to have a certain familiarity with the stimulus and the produced sound pressure level, and constant in time; this is because if the sound pressure varies with no displacement of the sound source, the sensation would be exactly the same at any variation of the distance.

Another phenomenon should also be considered (Mach, 1865): a given increase of the sound pressure level generated by the source (caused both by the source coming closer and by a variation in its intensity) corresponds also to a "darkening" (attenuation of high frequencies) of the sound, and *vice versa*. This can easily be explained through observing the Equal Loudness curve (see Figure 16 in Chapter 1), which shows that with a decrease in the pressure level, the perception of low frequencies drops much faster than the perception of high frequencies and *vice versa*, resulting in a decrease of the brightness of the sound source increasing the distance.

Other experiments into the link between the sound pressure level and the estimation of the distance outlined another significant phenomenon. Gardner (1969) carried out an experiment with five individuals and two loudspeakers, one placed at six metres and the other at nine metres from the listener. A speech signal was reproduced first from one, then from the other, speaker at different SPL. The sensation of the distance resulted as the same for both loudspeakers when the SPL at both ears was equal, and the 1/r law was not at all relevant. In fact, a reduction of 20 dB, instead of 6 dB, was interpreted as being a doubling of the distance of the auditory event.

In summary, when the sound pressure is the only attribute usable by the hearing system to estimate the distance of the sound source, as is exactly the case outlined in the first point, a strong discrepancy can be found between the distance of the sound event and that of the auditory event. Von Bekesy (1949) hypothesized that the auditory space has a limited extension, and the existence of a limit beyond which the distance of a sound source appears as being constant. This limit is commonly known as the auditory horizon.

Considering then the second point in the list, to the SPL reduction effect of frequency independent air absorption needs to be added the fact that the propagation medium (in this case, the air) is not perfectly linear, particularly for larger distances. Experiments carried out using low-pass filters at 10 kHz and 7 kHz demonstrated that with a lower cutting frequency, the sound source is perceived as being farther away (Coleman, 1968). In these experiments a phenomenon similar to the auditory horizon was reported. Furthermore, for a distance greater than 15 metres the familiarity with the sound source plays a very important role. For example, the sound of thunder is always perceived as being generated far away, even when produced by a loudspeaker located close to the listener (Aschoff, 1963).

The third point requires that the perception of the distance for close sound sources (a distance shorter than three metres) needs to be considered. The sound pressure generated by the sound source at the eardrum follows the 1/r law, but in this case the sound waves arriving at the head cannot be approximated as being plane; therefore, the distortions introduced into the signal by elements such as the head and the pinna can no longer be considered independent of the distance of the sound source.

The first experiments into the links between the HRTF and the distance of the sound source were carried out by von Bekesy in 1938. It is true that the transfer function of the external ear varies markedly at short distances, yet it must be asked whether the human hearing system is actually able to make any use of these changes for establishing the distance of the sound source. The work of Law (1972) shows how the spectral cues do not seem particularly important for the perception of distance. Pierce (1901) carried out experiments into the localization blur for distances between 150 and 50 cm, measuring values of Δr = 13-15 cm. Shutt (1898) proved that the precision of the distance estimation was much more accurate for broadband than for narrowband stimuli.

In conclusion, it can be said that HRTF spectral variations exist for sound sources close to the head: nevertheless, a threshold can be established; Blauert (1996:131) places it at 25 cm from the head. Farther than this, the spectral cues are far less intense or relevant in terms of distance estimation. It also needs to be taken into account that precise studies into and descriptions of the attributes influencing the perception of distance have not yet been performed.

Regarding the fourth and final point, the IHL effect will be described and analysed in the following section.

Section 5.3 then gives information about an experiment carried out within this research project about the ILD and spectral variations for close sound sources.

5.1.2 Inside the Head Locatedness

It is relatively common to perceive sounds as coming from a sound source located inside the head, such as when producing a sound with the mouth closed, for example. The IHL phenomenon can also occur for sound produced by sources located in the exterior space.

Generally, when presenting a diotic stimulus over headphones the sensation is of IHL. Inverting the phase for the signal at one ear only, the perceived source also moves towards the rear, yet still remains inside the head. The IHL can create various problematic issues in the context of binaural spatialization, for which the signals need to be reproduced over headphones. Various pieces of research have been accomplished on this topic:

- Kietz (1953) asserted that the IHL is to be attributed to natural resonances of the microphones of the dummy head and of the headphones.
- Frannsen (1960) attributed the IHL to an "overmodulation" of the nervous system.
- Schirmer (1966) combined different phenomena: the invariability of the signal to the movements of the head, the reaction of the eardrum to an impedance differing from that in free-field, the mechanical pressure of the headphones on the head, and the absence of sound presented to the rest of the body.
- Sone et al. (1968) suggested that the IHL could be due to an unnatural proportion between the signal delivered through the air and the signal delivered through bone conduction.

- Reichardt and Haustein (see Reichardt, 1968) carried out a series of experiments using hearing tubes (a type of tube with two openings inserted inside the ear, in order to bypass the pinna), and found that the IHL was nearly always present when the filtering effect of the external ear was absent. The conclusions coming from these studies established important prerequisites for the appearance of the IHL:
  - The signals at both ears need to be similar and coherent, as for a diotic stimulus.
  - Each of the two sound sources needs to be close to the ear, or at least perceived as close when the two signals are presented separately.
- Laws (1972) carried out different experiments in order to eliminate the IHL sensation. He placed two probe microphones in the ears of a listener. Through them he recorded two different stimuli: one generated by a loudspeaker placed in front of the listener at a distance of three metres, and the other generated by a pair of headphones (exactly the same signal generated by the same tension arrives at both speaker and headphones). The experiment was conducted with twelve different individuals. In the second stage, Laws created a filter to simulate the spectral difference between the two stimuli, averaged on the twelve subjects. The result was that a signal filtered and presented through headphones was localized outside the head, approaching closer and thus increasing its intensity. The distance perceived resulted in being shorter than the distance for the loudspeaker placed at three metres from the listener. This can be due to the averaging of the filter among the twelve different subjects.

In more recent studies, Weinrich (1992) investigated the elimination of the IHL when linked with the perception of frontal sound sources for headphone signals. Front-back discriminations seem linked mainly to the loudness parameter, and while the ILD can be considered the most important parameter for lateralization of sound sources, the ITD seems highly relevant to externalization. Weinrich also developed a filter called the Individual Frontal Perception Key (IFPK) composed of different notches typical of signals coming from frontal positions. Of course, the positions of the notches need to be individually calibrated according to various parameters, including the shape of the pinna and the dimensions of the head and torso.

Brookes et al. (2005) impute the IHL for binaural signals to the symmetries between the left and right HRTF channels. The symmetries are often the result of simplifications of

the HRTF filter; for example, the MIT Kemar HRTF set has been measured only for one half of the azimuth range, then replicated specularly for the second half.

Hartmann *et al*. (1996) understood and proved the importance of the ITD for the externalization of sound sources. Furthermore, they outlined that low-frequency phase components were much more important than high-frequency ones, also defining a boundary at 1 kHz; this may have a strong link with the duplex theory for the lateralization of sound source described in Chapter 3.

The link between binaural environmental simulation and the externalization of headphone signals has been investigated (see Begault, 2002). An accurate simulation of the surrounding environment, resulting in a surround reverb applied to the signal presented to the listener, results in an increase of the externalization of the perceived sound sources.

The IHL issue has been demonstrated as highly important to a discussion of binaural spatialization. It will, therefore, be returned to within this thesis, in Chapter 6 and Chapter 8.

5.2 The distance cues

Taking into account the above points, an analysis will now be carried out of the distance cues. The localization cues were described and analysed in Chapter 3 in the discussion of the binaural perception of the angle; here, information will be given about the parameters in the signal input into the hearing system used for the estimation of the distance of the sound source.

5.2.1 Attenuation of the air

When the source is far enough distant that the wavefront arriving at the subject's head can be considered plane, the sound pressure p_{rms} changes in a way inversely proportional to any change of the distance r, thus adhering to $1/r$. Where the intensity of the signal is concerned, it can be said that this decreases following $1/r^2$ (the inverse square law), therefore decreasing by 6 dB at each doubling of the distance.

These laws may be considered as valid in free-field situations; however, in closed environments, due to the reflections from the surfaces and walls, the reduction will be smaller with the increase of the distance and *vice versa*. In order to acquire a correct estimation of the air absorption in a closed environment, information would be needed

about the type and dimensions of the environment, the materials, the shape of the walls, and other factors. It could, for example, be estimated that in a closed environment the intensity of a signal produced by any sound source instead of decreasing by $1/r^2$, actually decreases by $1/r$, and that the pressure decreases by $1/\sqrt{r}$. Similar behaviours of intensity and pressure reduction are likely to be found in a long and low-ceilinged room. The estimations are simple; for a fully accurate calculation, as said before, more information is needed.

Regarding free-field sound sources located at a distance of greater than 15 metres, the phenomenon of air absorption can no longer, as stated in the previous sections, be considered flat in terms of frequency response. The distortion generated by the medium (in this case, the air) results in greater absorption for higher than for lower frequencies. This phenomenon can change perceptibly according to the speed of the wind and to the humidity and temperature of the air.

Table 1 reports estimations of intensity attenuations (in dB) for each 1000 metres distance at different frequencies and with different relative humidity (*see* Smith, 2009).

Relative Humidity	**Frequency in Hz**			
	1000	**2000**	**3000**	**4000**
40	5.6	16	30	105
50	5.6	12	26	90
60	5.6	12	24	73
70	5.6	12	22	63

Table 1. Typical values of air absorption for different frequencies and with different relative humidity (adapted from Moorer, 1979).

5.2.2 Direct-to-reflected signal ratio

The direct-to-reflected signal ratio is certainly not a parameter to be considered in free-field conditions. When signals are played back within a closed environment, though, the ratio between the direct signal and the reflected one may play an important role in the perception of the distance. It can therefore be stated that, because true free-field conditions are very rare, reverberation is an essential parameter within the distance cues. If the sound pressure of the direct signal decreases following the $1/r$ law, the same does not necessarily hold for the reflected signal. It may be assumed that in a small, enclosed environment the amplitude of the reverberant signal produced by a source at constant intensity yet at varying distances from the listener(s) varies little, while in a larger and

more open environment it varies more widely. Many parameters linked with the reverb are related to the environment where the signal is presented, such as reverb length, early reflection delay, and colour. An approximation can nevertheless be made about the direct-to-reflected signal ratio, considering that while the direct signal is attenuated following the *1/r* law, the reverb attenuation follows the *1/√r* law (*see* Chowning, 1971). The sum of the two formulae may thus provide an estimation of the air attenuation in a closed environment with a given signal and variations in the distance.

The importance of early reflections as related to the accuracy of localization, and an understanding of the environmental characteristics, such as the dimensions of the environment, distances from the walls, or materials of the reflecting surfaces, must be highlighted. The perception of early reflections must be considered a complex matter; as a simple example, the reflected sound may be thought of as being generated by mirror-image sound sources, therefore generating problems within the sound localization process.

Three different aspects will be outlined and briefly analysed:

- Which are the detection thresholds for early reflections?
- How much directional information is gathered from the early reflections?
- How much environmental information is gathered from the early reflections?

Beginning with the first point, detection thresholds have been measured for early reflections, determining that a particular reflection (the test reflection) is less audible if additional reflections are present between the primary sound and the test reflection itself (*see* Burtgorf, 1961; Seraphim, 1961 and 1963). With time, then, the reflections overlap to an increasing degree, and the detection of single reflections is no longer possible. The resulting time function can be in fact described only by means of statistical signal theory. Early reflections may therefore be consciously detected within a certain temporal distance from the direct sound, depending on their number and intensity.

To address the second point: in establishing the position of the auditory event, the auditory system takes into consideration (subject to certain conditions) coherent components of the ear input signals that arrive within up to one or two milliseconds after the first component (Blauert, 1996:272). Given the description in Section 3.6.3 of the Precedence Effect, the signals reaching the ears are interpreted as a single auditory event if the temporal distance between them is sufficiently small, in the order of ap-

proximately 4-5 ms for clicks and up to 40 ms for complex tones. The localization of the sound source is then performed through analyzing only the signal that arrives first; it allows the resolution of ambiguities when, for example, the level of the reflected signal, arriving from different directions, is higher than that of the direct signal. Thus, the issue of the directional information given by early reflection components becomes extremely complex. If in certain cases the "interaurally coherent" reflections arriving shortly after the direct signals can be analysed directionally by the hearing system, in other cases the position of the sound source is determined only by the direct signal component. Unfortunately, few studies exist into this aspect, therefore the question cannot be answered fully.

As regards the third and final question, the issue seem to become simpler. Even when the early reflections are not analysed directionally, the information they provide to the hearing system about the acoustic characteristics of the environment where the sound source and the listener are located can often be particularly relevant. If a quantity of information on the dimensions of the environment can be established by the human hearing system through its analysing the late reflections, which constitute the reverb components of the signal, information such as distance from the walls and disposition of other acoustic obstacles within the environment may be gathered from the delay and intensity of the first reflections, and from their overlapping parts.

It may be summarised that even if all the spatial hearing mechanisms have not yet been thoroughly investigated and understood, the importance of early reflections in sound localization and environmental acoustics awareness has often been underlined within various studies and research.

Different studies carried out at the NASA-Ames Research Centre by Durand Begault and his team (*see* Begault. 1992; 2001) repeatedly verified how early reflections are important for the externalization of sound sources (*see* also Section 5.1.2). In the context of binaural spatialization and specifically the simulation of three-dimensional soundfields over headphones, it is understandable how the simulation of early reflections and, more generally, of reverberation becomes extremely important in the obtaining of a realistic and precise spatial effect, and how sound sources spatialized using only anechoic HRIR would most probably sound as though they are located inside the head of the listener.

5.2.3 Spectral cues

Section 5.1.2 has already stated that it is not particularly clear whether the distance parameter can generate spectral differences within the HRTF. For larger distances of more than three metres, different researchers seem to agree on the fact that there are no spectral modifications of the HRTF, while for shorter distances the issue increases in complexity.

For sound sources closer than one metre, when the wavefront can no longer be approximated as plane, spectral differences can be found reducing the distance between the source and the ear. Considering for example a source located at 18 cm from the centre of the head, in a frontal position: when the same source is placed at a one-metre distance, the left and right ear azimuth discrepancies could be considered as irrelevant. In contrast, for a distance of 18 cm (the diametre of the head of a typical human being) the source will be located at -30° of azimuth for the left ear and 30° of azimuth for the right ear. In this case the position (azimuth) of the source itself, calculated from the centre of the head, is always 0°, independently of the distance. It is, therefore, self-evident that spectral differences in the HRTF exist between a source at one metre (approximately 0° azimuth for both ears) and at 18 cm (Δ 60° between the left and right angle). The author is currently carrying out studies into this in collaboration with LIMSI-CNRS, Orsay, France.[1]

For the same reasons, that is, the curvature of the wavefront, the shadowing effect of the head may exert different influences depending on the distance of the sound source. Therefore, the ILDs for close sound sources can change radically, thus varying the distance perceived. An experiment into this aspect has been performed during research towards this Ph.D., and the results are described in the following section.

5.3 ILD variations for close sound sources

Different researchers investigated the variations of ILDs for close sound sources (< 2 metres). ILDs are found to increase with the decrease of the distance, and *vice versa* (Brungart, 1999; Fukuda, 2003). Nevertheless, no clear information is available whether or not this phenomenon could be considered as being frequency dependent – the reason why the study has been carried out within this Ph.D. research.

1 *See* http://www.limsi.fr

An experiment was planned and performed into the measurement of the variations of Interaural Level Differences for close sound sources. HRIRs have been measured at 90° of azimuth and 0° of elevation for different distances, between 1 m and 20 cm, calculated from the centre of the head (the medial point of the axis that passes between the two ears).

It should be stated that the results and data emerging from the experiment have not yet been analysed in depth; the time constraints of the Ph.D. research project excluded it. Nevertheless, some first results have been outlined, and further analysis and research have been planned (*see* Chapter 10).

5.3.1 Hypothesis

Following what has been said in Section 5.2.3 in respect of the curvature of the wavefront, for small distances the ILDs should be dependent also on the distance of the source. Studies have been carried out by Duda *et al*. (1998) on near-field ILD analysis for an HRTF simulated from a spherical head model. They demonstrated that low-frequency ILD variations become greater with distance for ranges shorter than approximately five times the radius of the sphere (head). Other studies from Brungart (2001), while investigating the effect of the simulated distance on increasing the intelligibility of speech signals in competition in an auditory display applications, showed how the ILDs for HRTFs measured at different distances from a Kemar mannequin varied with distance, if this was indeed shorter than one metre.

The goal of the experiment was therefore to investigate how the ILDs vary with distances between one metre and 20 cm from the head. The experiments of Brungart were duplicated, using a particular "*dual concentric*" Tannoy loudspeaker (of which a closer description will be offered in subsequent sections), and the results are analysed more closely. The aim of the research by Brungart cited above was not to study and analyse the measured nearfield HRTFs; it was to employ them for an experiment in a speech competition.

5.3.2 Experiment set-up

The measuring system used for this experiment, and the room where it was conducted, are the same as those used for the experiments in the measurement of HRTF (*see* Chapter 4).

A preliminary experiment exposed problems in the use of a dual driver loudspeaker (the Genelec 8040). As shown in Figure 11 of Chapter 4, the acoustic axis of the Genelec loudspeaker is located precisely mid-way between the centres of the tweeter and of the woofer. While this is obviously an approximation, it should be considered valid when the distance between the speaker and the listener is greater than one metre. At shorter distances, the respective positions of the woofer and the tweeter have a much stronger effect. For example, if the loudspeaker is at 20 cm from the head at 90° of azimuth, the tweeter will have an elevation position of approximately 30° and the woofer of -30°; such displacement may generate substantial distortions during the experiment.

For this reason, a different loudspeaker was used in the current experiment: the Tannoy Century 801 passive loudspeaker (coupled with a Yamaha A100 power amplifier). The speaker is built with the proprietary Tannoy Dual-Concentric technology, placing the tweeter above the woofer in a central position. The centres of the tweeter and the woofer thus coincide. Therefore, the acoustic axis of the speaker can easily be determined and the effect of the relative displacement of the two drivers is eliminated.

5.3.3 Results

Figure 1 shows the different spectra of the ILDs for sources at different distances (at 90° azimuth and 0° elevation, and between 20 cm and 90cm in 10 cm increments), compared with the one-metre ILD curve.

It can easily be noticed that the ILDs change drastically over the distance range, becoming larger with each decrease in distance. For the first six distance increments, those between 90 cm and 40 cm, the ILDs for higher frequencies are greater than those for lower frequencies; however, for distances of 30 cm and 20 cm the ILDs of lower frequencies increase very rapidly, and at 20 cm the ILDs seem to be more or less homogeneous over the whole spectrum range.

The reason for this can be established when different phenomena are considered. A spectral difference in the ILDs at decreasing distances is expected in terms of a larger ILD for higher frequencies as compared with lower ones; this is due to the obstacle posed by the head, which absorbs a greater number of higher frequencies than lower ones for closer distances. The sudden increase in ILDs for lower frequencies at very close distances cannot as easily be explained. The proximity effect was initially exam-

ined as the cause, because despite of the fact that the microphones placed in the dummy head are omnidirectional, their positioning inside the dummy head itself makes them directionally sensitive (for further information on the proximity effect, *see* Martin, 2006). Nevertheless, the distance between the loudspeaker and the microphone, even at 20 cm distance, should be greater than 10 cm, thus the proximity effect cannot be considered the only cause for this increase in the low frequency components within the ILDs.

As an overall result, ILDs for higher frequencies seem to increase linearly with a decrease in distance, while for lower frequencies the increase seem to be exponential. Nevertheless, a more careful analysis of the signals and sonograms of the left and right channels is required in order to achieve a better understanding of those behaviours. Because these particular issues are not strictly related to the main area of this research, further analysis is postponed until after the completion of the Ph.D.

5.4 Brief summary

The present chapter introduced the binaural phenomena for the perception of distance. Starting with a description of the mechanisms involved in the estimation of the distance of a sound source, the Inside the Head Locatedness effect has been analysed, then three typologies of distance cues have been outlined and described.

In the last section, an experiment performed within the Ph.D. research on the ILD variations for close sound sources has been described, and a brief analysis of the results has been provided.

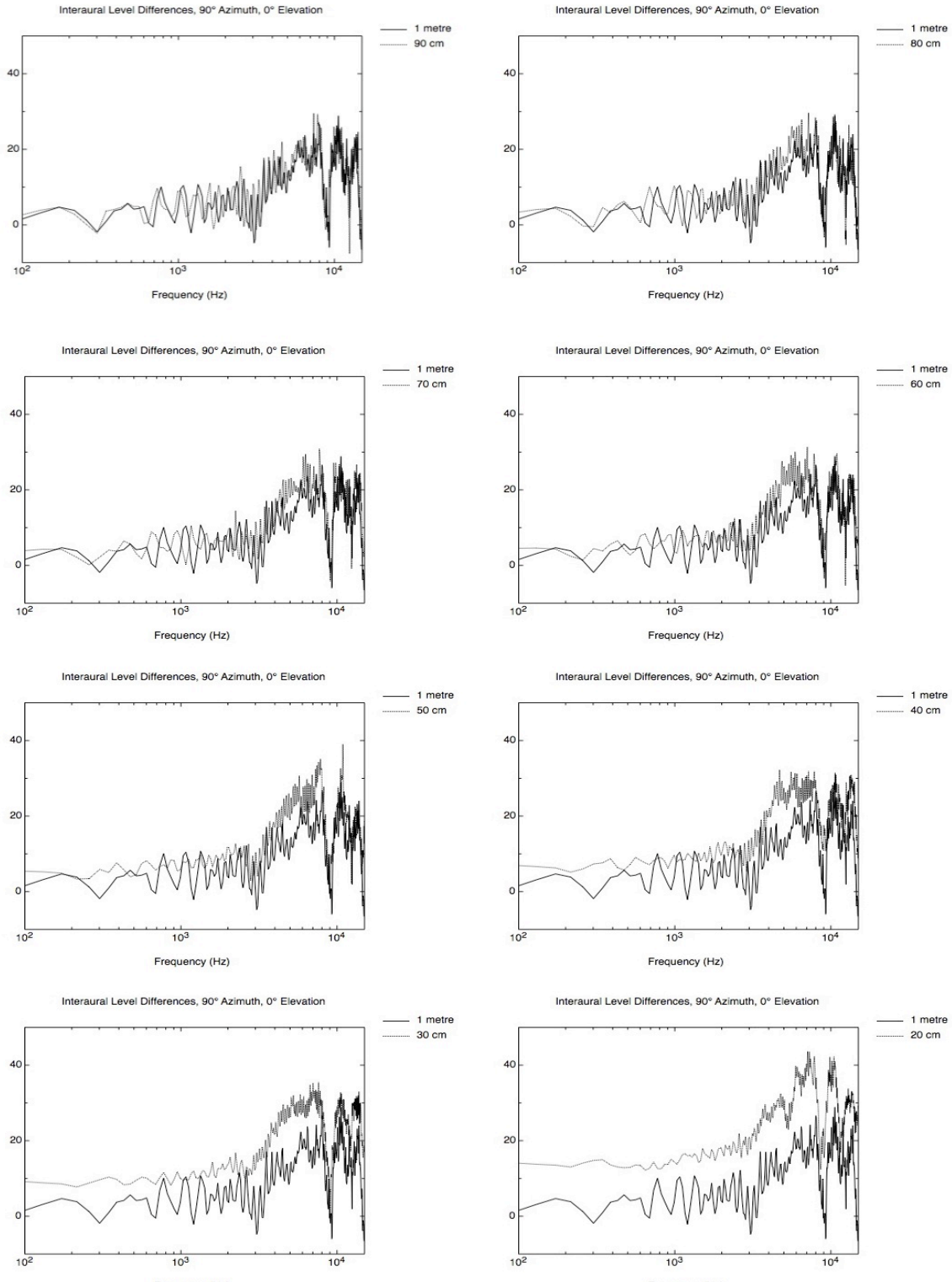

Figure 1. The spectra of the ILDs at different distances (90° azimuth and 0° elevation, and between 20 cm and 90cm in 10 cm increments), compared with the one-metre ILD curve.

Chapter 6

6. Distance Simulation and Binaural Reverb

The previous chapter, Chapter 5, has described the psycho-acoustic mechanisms for the estimation of the distance of a sound source; three distance cues, linear and non-linear components of air absorption, the direct-to-reflected signal ratio, and spectral cues have been detected and described. In the present chapter, the analysis of the three distance cues will be extended and a technique proposed for the simulation of the distance for a given virtual sound source over headphones, as well as for the simulation of different virtual acoustic environments. Chapters 5 and 6, centred on the auditory perception of distance, constitute a 'mirroring' of Chapters 3 and 4, which were centred on the perception of auditory angles.

After a brief overview of the state of the art in binaural distance and reverb simulation referred to in Chapter 2, experiments on the measurement of the BRIR or Binaural Room Impulse Response (*see* Chapter 1) will be described, and innovative methods for the binaural simulation of the distance of the sound source and of the acoustic environment will be presented and analysed. These methods are based on the manipulation of HRIRs and BRIRs, and form the main area of originality of this research; in the literature, no other references could be found to this aspect.

It is important to underline that binaural distance and reverb simulation are two distinct yet intimately related topics. The simulation of distance is linked to the simulation of the three distance cues, one of which has been identified as the direct-to-reflected signal ratio; in the simulation of this last cue is where the binaural reverb gains its high degree of significance. While for the simulation of the distance an arbitrary environmental acoustic simulation may be employed (as will be seen in Section 6.3, below), the development of a proper binaural reverb technique can add flexibility to the simulation of acoustic differences between distinct environmental configurations, and this is exactly what will be described in the following sections.

The two main techniques analysed within this chapter have then been tested through extensive perceptual listening experiments, information on which is given in Chapter 8.

6.1 State of the art

Chapter 2 provided a report compiled into the research on the state of the art in the sound spatialization field. Various algorithms, techniques, software and hardware systems were analysed and briefly evaluated of their performances in terms of sound spatialization quality and accuracy. At the end of the overview, in Section 2.7 of Chapter 2, final considerations were made of the different functions and characteristics of the enumerated systems; three of the points listed in that section need to be taken into consideration in justification of the work performed in the context of this current chapter:

- Localization of the apparent image of the sound sources outside the head. In none of the binaural spatialization algorithms that have been tested was it possible appropriately and precisely to perceive sound sources located outside the head. Even if sometimes the soundscape seemed to move from inside to outside the head of the listener, the positioning of the sound sources was everything but clear and precise, probably due to the fact that the reverb simulations were insufficiently accurate. A few of the several research groups have indeed specifically focused on this issue, also known as the "externalization of sound sources". AMES (NASA) has carried out extensive research into this topic, but unfortunately very little information is available on their research. While other publications have been written (Thomas, 1997; Weinrich, 1992; Hartmann, 1996; Takeuchi, 1998, and Brookes, 2005), no actual software or hardware implementations are available.
- Simulation of distance. There exist a few algorithms implementing the simulation of distance. Panorama (WaveArts) has the greatest number of control parameters, such as reverb and reflection, although the simulation is achieved only through loudness and the direct/reflected sound ratio. No systems could be found incorporating the simulation of spectral cues for the perception of distance. As regards the work of the research groups, only a few have focused on the variation of localization cues in the function of distance. Their research is often centred more on the perception of distance in reverberant fields, thus on the influence of all of the acoustic parameters of the room, than on the individual analysis of the localization cues for the perception of distance. No HRIR database with distance parameter variations was found; all of the most important (MIT and CIPIC) are measured at a fixed distance from the head.

- Binaural reverb. It was difficult to find systems implementing reverberation in the binaural domain. Panorama (WaveArts) offers a greater quantity of parameters in terms of reverberation, yet from the information that could be gathered it is a simple binaurally enhanced stereo reverb. The Beyerdynamic binaural simulation offers a reverb, although its generation is limited to a 5.1 set-up. Both the IRCAM (SPAT) and the IEM (Graz) research groups developed binaural reverbs based on multichannel (mainly Ambisonics) rendering. Very few binaural reverb algorithms based on the BRIR (Binaural Room Impulse Response) could be found, and for these the flexibility of the environmental simulation was limited simply to the room in which the BRIR was measured.
- It is clear that the second and third points are strictly linked to that which will follow within this chapter (binaural distance and reverb simulation), whilst for the first point further explanations are probably needed, therefore Section 5.1.2 outlined links between the externalization of binaurally simulated sound sources and an environmental acoustic simulation (see also Begault, 2002). It has been clearly demonstrated how the externalization of sound sources is linked with the simulation of distance within a real non-freefield environment, given that sound sources are localized inside the head when their perceived distance is smaller than the radius of the head itself, and is therefore linked with the simulation of various environmental acoustic cues.

The state-of-the-art research outlined how a proper binaural simulation of distance and of environmental acoustic cues is missing within the consumer and professional systems specified; this is also the case within the listed research groups, resulting in less effective, less realistic binaural spatialization performances and in the appearance of the IHL effect. The attempt of the research stage described within the present chapter is precisely to fill this gap, and to implement within the binaural spatialization algorithm a proper section for the simulation of distance and of the reverb.

6.2 The measurement of HRIRs and BRIRs

The starting point of this stage of the Ph.D. research needs to be found in the extension of the available HRIR and HRTF material. Referring to Chapter 4, the measured HRIR database is anechoic or pseudo-anechoic depending on the IR editing procedures (see

Section 4.4.2). In order to perform a proper binaural distance and environmental simulation, other HRIR databases were measured in non-anechoic environments.

In this section, information will be given on the HRIR and BRIR (Binaural Room Impulse Response, see Chapter 1) measurement sessions performed in different environments, as well as on the editing of the measured IR. The way these materials are then used for the simulation of distance and of the reverb will be described in Sections 6.3 and 6.4.

All of the IR measurements were carried out according to the technique, procedures, measuring systems, and calibrations already described in Chapter 4.

6.2.1 HRIRs at different distances

The first addition to the HRIR database already measured (see Chapter 4) has been an HRIR measurement experiment for distances greater than one metre. Section 5.2.3 reports research demonstrating how spectral cues seem irrelevant for sources at more than three metres' distance. However, it is also true that researchers do not fully agree on this, and that further studies are still needed in order appropriately to understand the HRTF spectral differences when varying the distance of the measured source. For example, while for frontal sound sources no spectral differences are found at distances of between 0.5 and 4 metres, for lateral sources (at 60° of azimuth) spectral cues seem to become important for the estimation of the distance, and the accuracy of the estimation itself is greater (Nishimura, 2004).

For this reason, it has been considered important to measure the HRIR for distances greater than one metre in the same room where the one-metre HRIRs have been measured (see Section 4.3.1). The positions of the measured HRIRs are given here in Table 1:

DISTANCE	ELEVATION	AZIMUTH	SAMPLED POSITIONS
2 metres	-20°, 0°, 20° and 40°	Fixed steps, each 10°	144
4 metres	-20°, 0° and 20°	Fixed steps, each 10°	108
6 metres	-20° and 20°	0°, 90°, 180° and 270°	8
	0°	Fixed steps, each 10°	36
9 metres	-20° and 20°	0°, 90°, 180° and 270°	8
	0°	Fixed steps, each 10°	36

Table 1. Recapitulation of the measurement positions of the HRIRs with azimuth and elevation increments at different distances.

The reason for the different azimuth scales for -20° and 20° of elevation for distances of six and nine metres can be explained through considering the dimensions of the "Diffusion Room". At a distance of six metres and 20° of elevation, the loudspeaker for the reproduction of the sweep signal would have needed to be placed at 1.65*[6sin(20°)] =~ 3.7 metres of height, i.e., far greater than the actual height of the room itself (and for a distance of nine metres, the height would have been concomitantly greater). The solution applied to the experiments has thus been to incline the dummy head and to leave the loudspeaker at its same height. Because the HRIRs were edited in order to remove the reflections of the sound signals (including the reflections from the floor), the result was similar to having varied the height of the speaker itself. In order to simplify the measurement process, at this moment it has been considered sufficient to measure for these elevation angles the HRIRs at each 90° of azimuth, and not at each 10° as was the case for 0° of elevation.

These HRIRs were edited in order to isolate the direct signal path components, already described in Section 4.4. Another factor to be considered for future reference is that the linear air attenuation cue for these measurements has been eliminated, to achieve the more flexible management of this cue during the post-processing phase (see Section 6.3). All of the HRIRs measured at different distances were all in fact calibrated to maintain a maximum value of -6 dBfs (see Section 4.2.3). The increasing air attenuation

effect for larger distances has therefore been compensated by a higher amplification level for the loudspeaker reproducing the sweep signal.

6.2.2 BRIRs

The choice of the room for the measurement of the HRIRs described in Chapter 4 followed mainly the need for an environment as anechoic as possible, both to avoid wave reflections on the walls, ceiling and floor and to isolate better the HRTF components. In this case, for the measurement of a BRIR database the choice of the room needs to be determined according to the different characteristics of the environment where the measurements are performed, in order to achieve a clear variety in the acoustic characteristics of the different environments. The measurement experiments were therefore carried out in three different environments:

- Chantry House (medium-sized environment): a simple room with a wooden ceiling and tiled floor, its dimensions are 6 x 7 x 3 metres, and the RT60 is 0.6 sec (+/- 0.2 sec across the whole frequency range). In this room, HRIRs were measured at a distance of two metres for -30°, 0° and 30° of elevation.
- IOCT (medium-large environment): a large room with a highly reflective floor and ceiling, and with a pillar in the middle of the room itself; its dimensions are 11 x 8 x 4.5 metres, and the RT60 is 0.9 sec (+/- 0.2 sec across the whole frequency range). In this room, HRIRs were measured at a distance of three metres for 0° and 15° of elevation. Another measurement has been carried out placing the loudspeaker on one side of the pillar, and the dummy head on the other, in order to eliminate the presence of the direct signal, and therefore to isolate the effect of the wave reflections on the environment.
- Trinity Chapel (large environment): a medium T-shaped church with a wooden floor and ceiling, and partially wooden walls; the length of the T arms is 15 metres, their width is 8 metres, and the height is 6 metres. The RT 60 is 1.9 sec (+/- 0.3 sec across the whole frequency range). In this case the HRIR was measured only at a distance of one metre and at 0° of elevation.

6.2.3 Early reflection HRIRs

Another BRIR database was measured in the “drum room” of the Courtyard recording studio within the MTI undergraduate facilities, with absorbent materials on the walls,

ceiling and floor. The environment is a triangle with sides of 2 x 2 x 3 metres; it is 2.5 metres in height, and the RT60 is 0.1 sec (+/- 0.1 sec across the whole frequency range). In this specific case, in order to simulate an environment with many early reflections the walls were covered with acoustic reflective materials, changing the RT60 from 0.1 sec to 0.3 sec, and increasing the sound energy between 2 and 10 ms after the direct signal impulse. Figure 1 shows how the early reflections (ms 0.4) are nearly as strong as the direct signal (ms 0.2). The HRIRs have been measured at a distance of one metre and at -30°, 0°, 30° and 60° of elevation.

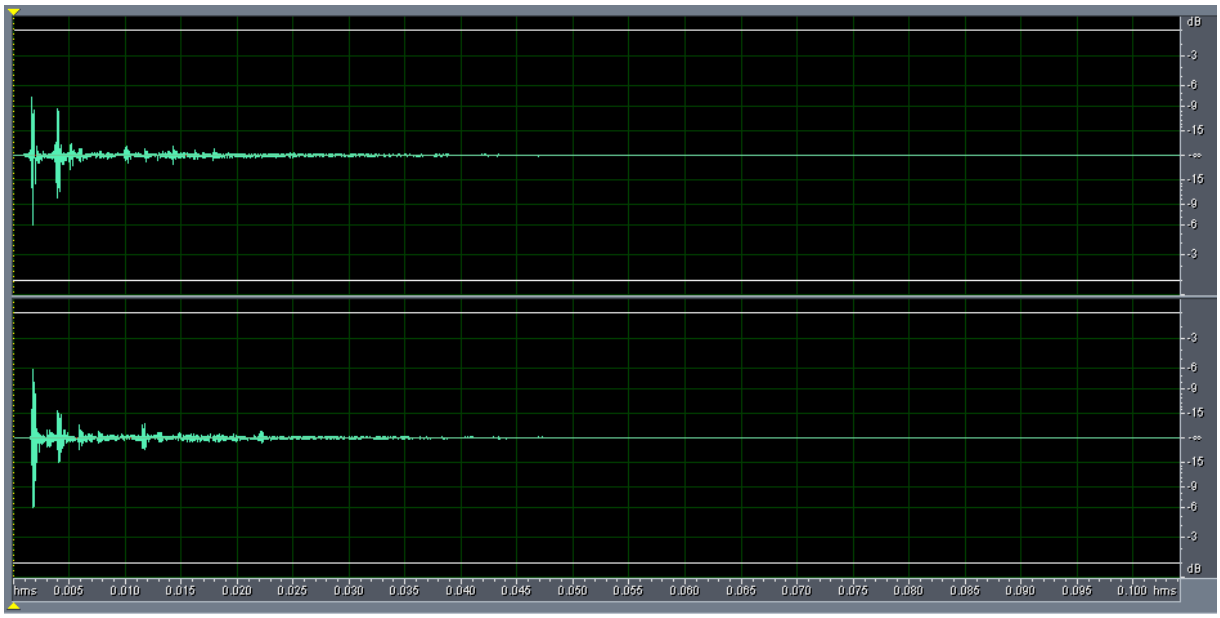

Figure 1. The oscillogram of the left and right channels from the Early Reflections HRIR for 0° Az and 0° El.

6.2.4 The organization of HRIRs and BRIRs

All of the measured HRIRs and BRIRs were processed and organized into a database, divided into folders according to the room in which each was measured, and following a naming standard similar to that used for the first HRIR database (*see* Section 4.4.2). Thus, an example is *Chantry_0El_0Az.aif*, where *Chantry* stands for the Chantry House BRIRs, *IOCT* for the IOCT BRIRs, *Trin* for the Trinity Chapel BRIRs, and *EarlRef* for the Early Reflections HRIRs.

The positions and typologies of the different HRIRs and BRIRs measured can be found in Table 2:

ENVIRONMENT	DISTANCE	ELEVATION	AZIMUTH	SAMPLED POSITIONS
Chantry House	2 metre	-30°, 0° and 30°	Fixed steps, each 10°	108
IOCT	3 metres + pillar	0° and 15°	Fixed steps, each 10°	108
Trinity Chapel	1 metre	0°	Fixed steps, each 10°	36
Early Reflections	1 metre	-30°, 0°, 30° and 60°	Fixed steps, each 10°	144

Table 2. Recapitulation of the measurement positions of the HRIRs and BRIRs for the different positions and azimuth and elevation increments at different distances.

6.3 Simulation of the distance cues

Three distance cues have been identified and described in Chapter 5. They are the parameters of the signal input into the ears used by the human hearing system in estimating the distance of a sound source.

In order to simulate the distance of a virtual sound source within a 3D audio rendering system over headphones (a binaural system), one that gives the impression of a sound source positioned outside the head of the listener, the distance cues need to be re-created within the processed signals after considering the psycho-acoustic mechanisms for the estimation of distance described in Chapter 5.

A first approach to achieving this would be to sample the distance around the listener's head in fixed or variable steps, and to measure HRIRs at the different distances. Considering for example the simulation of the distance between 0 and 30 metres, with the aim of sampling the distance every 50 centimetres, HRIRs need to be measured in spheres with a radius of between one metre and 30 metres, with a 50-centimetre step. If the same azimuth and elevation sampling scale are to be retained for all of the distances, as in Table 2 of Chapter 4, a total of 1,729 HRIRs need to be measured for each distance, resulting in a database of 102,011 HRIRs – far too large to be considered flexible and portable. Even if the simulation to 0° of elevation were limited, the resulting number of measured HRIRs would be 8,496.

A further issue needs to be considered; given that the distance cue is related to the direct-reflected signal ratio, the HRIRs used for the simulation would need to be measured in a semi-reverberant room (for more information on distance estimation performances in free-field and reverberant environments, *see* Zahoric, 2002), resulting

in a much larger quantity of HRIRs measured. A database of 8,496 reverberant HRIRs can easily assume the dimensions of several Gb of data.

A second approach, and that applied within this research, is to modify then combine different parts of relatively few HRIRs to achieve a more portable and flexible binaural simulation system. Distance cues are therefore generated and modified through the editing and processing of the HRTFs measured. To simplify the situation, the distance simulations are performed only on the horizontal plane. Further developments of the algorithm are planned (*see* Chapter 10).

Taking into account the information contained in Section 5.2, the following sub-sections will present and analyse methods for the simulation of the three localization cues.

6.3.1 Attenuation of the air

The inverse square law (*see* Section 5.2.1) states that the intensity of the signal generated by a source then varying the distance would decrease by -6 dB each doubling of the distance. As was observed in Chapter 5, this rule is valid in free-field conditions, i.e., when the environment does not interact with the reproduced signals (as in an anechoic chamber). It is true that the law is also valid if the signal under consideration is the only one to reach the listener along a direct path between the source and the listener him/herself. The attenuation of the air may be simulated with a simple gain reduction line that attenuates the direct signal, the anechoic HRIR (as will be seen in Section 6.3.2), of -6 dB at each doubling of the distance.

As stated in Section 5.2.1, for distances larger than 15 metres, the phenomenon of air attenuation can no longer be considered as flat in terms of frequency response. The distortion generated by the medium (in this case, the air) will result in a greater absorption for higher frequencies than for lower ones. Because this phenomenon is heavily influenced by many variables, including the temperature and relative humidity of the air, approximations and generalizations need to be performed for the simulation of this cue. For more information on the data used for these approximations, see Moorer (1979). Starting from the simulation of a distance of 10 metres, a low-pass filter is implemented with variable cut-off frequency and Q. The cut-off frequency varies logarithmically between 20 kHz (no audible difference) at 10 metres, 14 kHz at 15 metres and 6 kHz at 30 metres; the slope of the filter varies from 6 dB/Oct at 10 metres,

12 dB/Oct at 15 metres and 18 dB/Oct at 30 metres (detailed information about the implementation of this filter will be given in Chapter 7). For the simulation of sources at a distance greater than 30 metres, the phenomenon of the *auditory horizon* has been considered valid (*see* Section 5.1.1, and von Bekesy, 1949).

6.3.2 Direct-to-reflected signal ratio

The main difficulty linked with the simulation of the direct-to-reflected signal ratio cue is that it is impossible to generate the reverb and the early reflections using standard algorithms for stereophonic reverberation. These artificial reverberators work on a mono-dimensional soundscape simulation, in which the signal reflections can arrive only from the right or from the left of the listener's head. In the case of a 3D soundscape simulation (as occurs in binaural spatialization), the spatial limitations of these stereophonic algorithms would indisputably create problems both for the performance of the simulation and for the localization of apparent sound sources inside the head of the listener. As was observed in Chapter 2, many of the binaural spatialization algorithms available on the professional and consumer markets implement a reverb simulation using standard stereo reverbs, resulting in a decrease in the quality of spatialization and also in the appearance of the IHL effect. A perceptual test comparing a binaural reverb simulation with a stereophonic was performed and described in Chapter 8.

For this reason, an innovative environmental simulation is presented in the subsequent paragraphs, which will focus on the simulation of the second distance cues, while Section 6.4 will focus on the characterization of the BRIRs measured.

The difference between HRIRs and BRIRs is supplied by the environment in which the IRs are measured. An HRIR is typically measured in an anechoic environment and it represents the response to a direct signal, while a BRIR is typically measured in a reverberant or semi-reverberant room. For this research, pseudo-anechoic HRIRs have been measured and edited, as well as BRIRs measured in different rooms.

Through the processing and editing of each of the measured IRs, the following sets have been created:

- Direct Path HRIRs (anechoic): impulse responses measured in a semi-anechoic environment. For this set, the HRIRs measured and edited as described in Section 4.4 have been applied.

- Early Reflections and Reverberant BRIRs: using the BRIRs measured in Trinity Chapel (*see* Section 6.2.2), the IRs have been split in order to isolate the direct path, early reflections, and reverberant components. The separation between the first two can be made only if the first reflection comes after the HRIR has decayed to a small value. Considering the work done by Sontacchi *et al*. (2002), the relevant (in this context, to sound source localization processes) spectral and temporal data in an anechoic HRIR can be found within the first 256 samples of the IR (sampled at 44.1 kHz). In the case of a first reflection arriving after longer than 1 ms (441 samples), this problem may be considered as resolved. Figure 2 shows how these components are clearly noticeable within the oscillogram of the IRs. The direct path components have been selected at between 0 and 5 ms from the beginning of the IR, the early reflection components between 5 and 15 ms, and the reverberant components from between 15 ms and the end of the IR. Of course, both the early reflections and the reverberant components have been faded in and out with a ten-sample window, given that the direct path IR only has been faded out; furthermore, it was zero-padded at the beginning in order to maintain the same temporal sequential order. The early reflections BRIRs have been padded with 5 ms times zeroes, and the reverberant with 15 ms. Due to the fact that the direct path components are taken from a different HRIR set (see the previous point, above), two BRIR sets were created through editing the measured Trinity Chapel BRIRs. They are an Early Reflections set and a Reverberant set; the direct path components have been eliminated.

The simulation of this second distance cue can be therefore performed through convolving the signal generated by the sound source with the three IR sets (direct, early reflections, and reverberant components), then performing a weighted mix among the three signals. At a distance of one metre, all components are summed at the same level, assuming that the ratio between these has not been rescaled during the measurement and the calibration processes, thus the levels have been kept the same as the ones of the room where the measurement took place. The three components are then scaled as follows (for more information about the principles behind these scaling values, *see* Chowning, 1971).

- The amplitude of the direct path component is proportional to $1/r$, which corresponds to the inverse law for the sound pressure level with increase in distance, and allows the level to be left unvaried at distance 1, i.e., at one metre.
- The amplitude of the early reflections component is proportional to $1/r^2$. The quantity of detectable early reflections actually decreases with distance more than does the direct signal. If the distance is increased, part of the early reflections becomes indistinguishable from the reverb components. The formula allows the level to be left unvaried at distance 1.
- The amplitude of the reverberant component is proportional to $1/\sqrt{r}$, therefore it decreases more slowly than the direct component at increasing distance. The amount of reverb in the signal at the ears does not vary perceptibly with distance, or at least far less so than the direct component. This formula, too, allows the level to be left unvaried at distance 1.

Another element with relevant importance when the weights of the different signal components are varied is the delay between them; this is true most of all between the direct path and early reflections components. As has already been said in this Chapter, the HRIRs and BRIRs used for the spatialization process were all measured at a distance of 1 metre; during the editing process silence (zeroes) were added before the early reflections and reverberant components in order to maintain the same temporal sequence of the original signals.

Therefore, when the distance to be simulated is 1 metre, no further delay is added within the processing. Furthermore, when greater distances need to be simulated, the early reflections pre-delay would become smaller, varying inversely proportionally with the distance. In this case, the early reflections components need to be anticipated slightly, in order to move them towards the direct path components and thus to simulate larger distances. The pre-delay is therefore varied from 0 (leaving the same spatial sequencing of the BRIR as measured at 1 metre) to -3.5 ms (sample accurate), placing the early reflections at only 1.5 ms after the direct path component. No pre-delay is applied to the reverberant components (again, for more information about the principles behind these values, *see* Chowning, 1971 and Shroeder, 1962).

In order to make the simulation more flexible, the final version of the binaural spatialization algorithm will allow the user to manage the pre-delay parameter irrespective of

the distance to be simulated. For further information on the actual implementation of the algorithm, *see* Chapter 7.

Of course, this is an approximation and generalization of the reduction values and pre-delays for the direct, early reflection and reverberant components (*see* Section 5.2.2). Future studies will permit finer calculations of these parameters in order to simulate different environment typologies (*see* Chapter 10).

It could be noted that in Section 6.2 two particular BRIR databases have been measured: one is early reflections, as was measured in the treated Courtyard recording studio, and the other is reverberant, with no direct path components. The latter was measured in the IOCT, placing the loudspeaker and the dummy head on either side of a pillar. The original plan was to use these as the early reflections (the former) and reverberant BRIR sets; it has been considered more appropriate to divide a BRIR into its different components, in order to limit the variables within the simulation. The variables include the different acoustic characteristics of the two environments where the BRIRs have been measured. Dividing the BRIRs thus allows more coherent BRIR early reflections and reverberant sets, measured in the same environment.

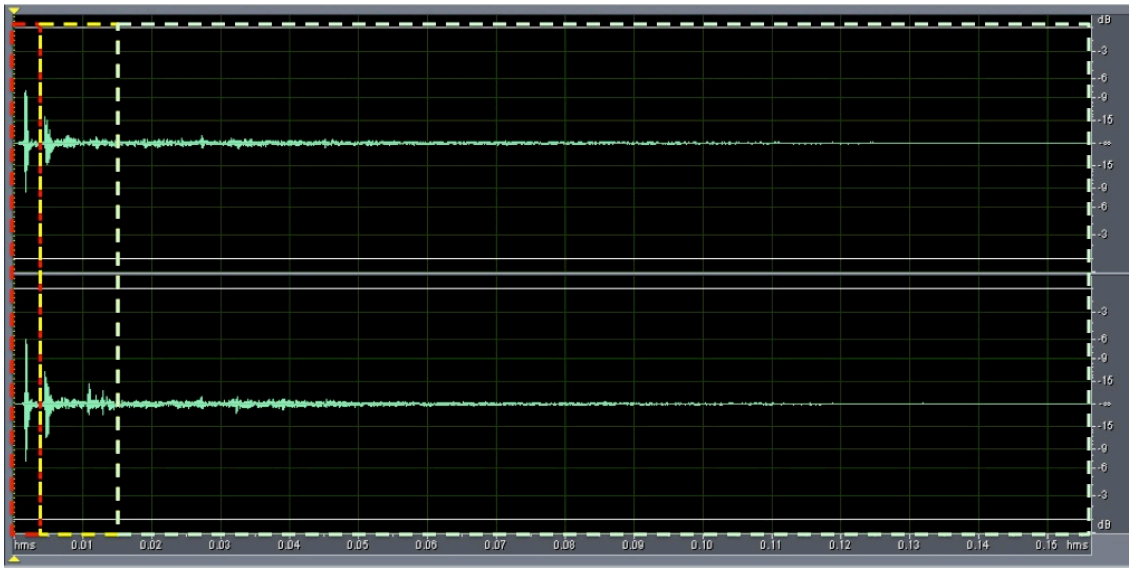

Figure 2. The oscillogram of the left and right channels from the Trinity Chapel HRIR for 0° Az and 0° El. Within the IR, it is clearly possible to differentiate the direct path components (outlined with a red dashed rectangle), the early reflections components (outlined with a yellow dashed rectangle) and the reverberant components (outlined with a green dashed rectangle).

6.3.3 Spectral cues

The last distance cues to be simulated are the spectral variations between HRIRs measured at different distances (*see* Section 6.2.1). In this case, the direct path signal component is composed through performing an interpolation of the source signal convolved with the correspondent HRIR, i.e., for the angles of azimuth and elevation, measured at different distances. For example, for the simulation of a distance of three metres, the source signal has been convolved with the HRIR measured at two metres and the HRIR measured at four metres, each summed with a -6 dB reduction, in order to perform appropriately a linear interpolation between the HRIRs measured at the two positions. For the simulation of sources at distances larger than nine metres, no interpolation is carried out, and the direct path signal component is retained as it is.

It should be stated that in the first version of the binaural spatialization algorithm (*see* Chapter 7), this simulation was not performed; therefore, only the one-metre HRIRs were used as direct path components, irrespective of the distance to be simulated.

6.3.4 Distance simulation summary and MaxMSP implementation

To summarise the information provided in this section: the source signal to be binaurally spatialized is processed in the following manner:

- The direct path signal component is created convolving the signal with the correspondent HRIR (azimuth and elevation) from the direct HRIR set, performing an interpolation between the HRIRs measured at different distances in order appropriately to simulate the required distance parameter.
- The early reflections and reverberant components were created through convolving the signal with the correspondent BRIRs (azimuth and elevation) from the early reflections and reverberant BRIRs set.
- The three components were then combined with appropriate weightings to achieve a simulation of the required distance.
- The signal output from this process is then input into a gain reduction line and into an equalization filter which has already been calibrated in order to simulate the required distance, in order to simulate the frequency-dependent and -independent components of the air attenuation cue.

A preliminary software prototype was implemented to verify and calibrate the different values and parameters for the simulation of distance. The implementation was achieved using the MaxMSP visual programming language (a copy of the developed application can be found in the CD attached to this dissertation, *see* Appendix E). A screenshot of the implemented software is shown in Figure 3.

This particular technique was then tested through an extensive perceptual experiment, on which a report can be found in Chapter 8.

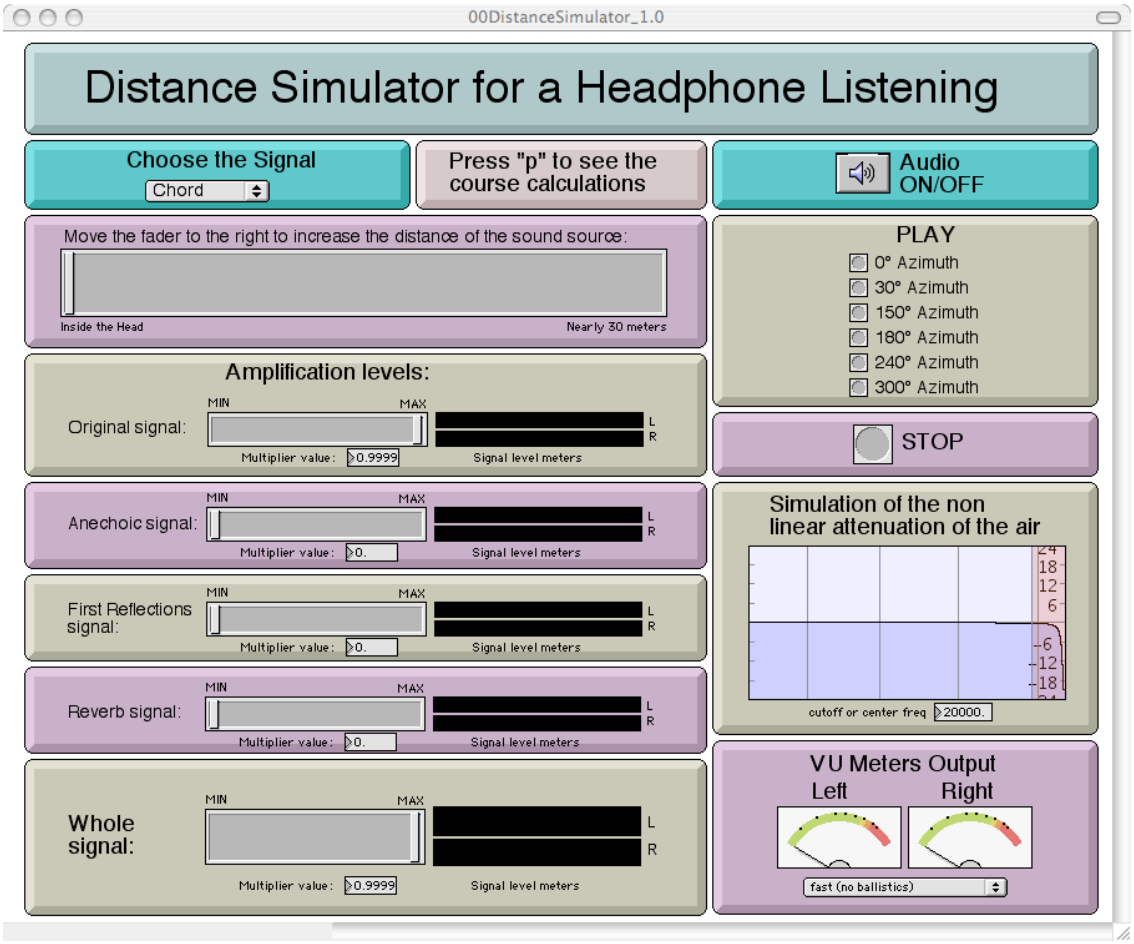

Figure 3. The MaxMSP platform used for the implementation of the prototyping of the distance simulation technique

6.4 The characterization of BRIRs

Taking the information outlined in Section 6.3 and given the fact that part of the ability of the human hearing system to estimate the distance of a given sound source is linked with the direct-to-reflected signal ratio and, further, that in order to simulate this distance cue HRIRs need to be measured in reverberant or semi-reverberant environments,

it may easily be remarked that the simulations performed may differ widely according to the acoustic characteristics of the environment in which the BRIRs are measured.

As before, the BRIRs measured in the IOCT serve as an example. There, the simulation would carry the acoustic characteristics of the IOCT environment, and the virtual sound source will appear as being located inside that specific place. A high quality standard stereophonic reverb simulator allows the user to establish the parameters of the characteristics of the simulated environment, such as the room dimensions, the damping coefficients for different frequencies, and the absorbent coefficients of the walls, ceiling and floor. It is impossible to achieve such flexibility when convolution reverbs are under discussion; these are based on the mathematical operation between the signal to spatialize and the IR of the environment to be simulated. Here, as in the case of the distance simulator presented in the previous section, the IR corresponds to that specific environment. To change within the simulation, for example, the dimensions of the room, another IR needs to be measured, taken from a different environment.

Nevertheless, some of the best known and most commonly applied convolution reverbs (for example, the Logic Audio Space Designer [1]and/or Audio Ease Altiverb[2]) allow the modification of the IR in, for example, simulating different sound colorations within the same environment. If a shorter reverb time is to be simulated for frequencies above 2 kHz, resulting in a 'darkening' of the reverb sound, the IR can be divided into frequency bands. Then, the bands above 2 kHz may be faded out more quickly than those below 2 kHz. Usually, these processes are performed on 6-8-band parametric equalizers, changing simply the fade-out time and curve, and allowing both establishing of parameters and the modification of the reverb time in limited frequency bands.

The environmental simulation technique presented in the previous section can be considered as a binaural version of a convolution reverberation, i.e., convolution with BRIRs and not simply with IRs; thus, a method similar to that described in the previous paragraph could be attempted in order to add flexibility to the simulated environment.

It must be emphasised that such processing has never before been attempted with BRIRs; therefore, this represents an actual innovation introduced by the research documented here.

1 *See* http://www.apple.com

2 *See* http://www.audioease.com

6.4.1 Cross-synthesis

Typically, crossed source/filter synthesis is performed through multiplying the FFT spectrum of a sound with the analysis-generated spectral envelope of another sound. The first sound is filtered using the second sound's spectral envelope. Regarding the sound acting as a filter, its spectral envelope can be analysed using different techniques, the simplest and more direct of which is the FFT analysis in a process usually known as "*generalized cross-synthesis*" (Lithaud, 2003).

Take as the envelope signal a stereophonic IR generated with the characteristics of the environment required to be simulated, which may be achieved by using, for example, an acoustic simulation software or a standard stereophonic reverberator, along with the BRIR (its azimuth and elevation corresponding with the position to be simulated) as the signal to be filtered; these are measured in a room with a sufficiently long and flat (in terms of frequency response) reverb time, which must be longer than that of the envelope signal: a cross-synthesis processing with these two signals would generate a BRIR with the acoustic colouring of the IR used as the envelope. The process may be considered an advance from the simple multiband parametric filtering described previously and performed by standard stereophonic reverberators; in fact, such a cross-synthesis process may be viewed as a variable gain and Q parametric filter, where the frequency resolution is given by the FFT analysis parameters. Further information on FFT analysis appears in Chapter 1.

The cross-synthesis process can be performed using various data emerging from the FFT analysis of the envelope signal: phase, amplitude, or both. Because phase information within the BRIR to be filtered is highly significant in terms of spatial hearing and binaural spatialization (for example, the ITDs), the cross-synthesis process is performed through considering only the amplitude data. Another factor that may be incorporated in order to preserve the binaural data coded in the BRIR is that a multiplier (< 1) may be used within the cross-synthesis process. Weighting can thus be given to the whole process; the proper multiplication coefficient may be chosen according to the simulation to be performed. Further discussion of this aspect appears in Section 6.4.2.

It would be virtually possible binaurally to simulate every kind of environment if the starting point were a sufficiently long BRIR, with few reverb time variations across frequencies (for example, the BRIRs measured in the Trinity Chapel, with an RT60 of

1.9 seconds +- 0.3 sec over the whole frequency range), filtered by the cross-synthesis method and using an IR generated with the required parameters including the dimensions of the room, and the materials of the walls.

The method nevertheless adds the flexibility of a standard stereophonic or multichannel algorithmic reverb to a binaural environmental simulation process based on BRIRs. Chapter 8 describes a perceptual test performed in order to verify the quality of such a process. In this case, the Trinity Chapel BRIRs were used as starting material, and three IRs were generated using the CATTAcoustic[3] simulation software. Three IRs were generated from the model of three different environments with different dimensions and acoustic characteristics, then used as spectral envelope filters performing cross-synthesis with the original BRIRs, allowing the binaural simulation of the three environments; further information appears in Chapter 8.

The implementation of this particular process has been accomplished through using Audio Sculpt 2.8[4]. With the built-in "Generalized Cross Synthesis" processing function, the Trinity Chapel BRIRs (edited in order to remove the direct path and the early reflection components) were used as the first signal, while different IRs generated with a CATTAcoustic model. It was chosen because this particular acoustic modelling software allows the creation of stereophonic IRs of only the reflected signal; in the case of the experiments in this thesis, in order to remove the direct path and early reflections components, the IRs were generated considering only the reflections beyond the Second Order for the reverberant component, and only the First Order for the early reflections component. It can be argued that in certain situations Second Order reflections can arrive before the First Order ones; nevertheless, due the architectural characteristics of the simulated environments, this does not happen in these particular simulations. The results were used as the spectral envelope signals.

The parameters chosen for the processing are the following:

- Cross Synthesis Mode: Constant Cross
- Sound 1 Amplitude Scaling: 1.0
- Sound 1 Phase Scaling: 1.0
- Sound 1 FFT Window Size: 2048 samples

3 *See* http://www.catt.se

4 *See* http://forumnet.ircam.fr/691.html

- Sound 1 FFT Oversampling: 12.5 per cent of the window size
- Sound 1 FFT Window Type: Hanning
- Sound 2 Amplitude Scaling: 0.5 – this value has been found after different attempts (all followed by an informal perceptual test). It represents the weight of the whole cross-correlation process, therefore the extent to which the first sound is modified by the second; 0.5 seemed to allow a suitable spectral envelope characterization of the BRIR without altering the localization cues embedded in the IR
- Sound 2 Phase Scaling: 0 – in this case, only the phase information of the BRIR has been maintained in order not to alter the localization cues, especially the ITDs
- Sound 2 FFT Window Size: 2048 samples
- Sound 2 FFT Oversampling: Follow Sound 1
- Sound 2 FFT Window Type: Hanning
- Q Factor: 1.0 – the Audio Sculpt default value has been left
- FFT Synthesis Oversampling: 1x – same window size as for the analysis
- FFT Synthesis Window Type: Hanning – same window type as for the analysis
- FFT Synthesis Gain: 6 dB – in order to compensate the 0.5 amplitude scaling for the second signal.

6.4.2 Possible questions and problems

A possible objection could be made that relevant spectral and phase information could be lost or highly altered through performing the cross-synthesis process on BRIRs. It would result in an alteration of the spatial perception of the spatialized signal, and therefore in a worsening in terms the quality of spatialization. Such a side effect should be previewed before performing the process. For this reason a perceptual test was carried out after the development of the spatialization algorithm. The results of the test seem to prove that the alterations (in terms of localization and distance cues) generated by the cross-synthesis process (scaling the spectral envelope signal with a 0.5 coefficient) are imperceptible, and that the modified BRIRs allow a quality of spatialization higher than in other binaural acoustic environment simulation techniques. When processing the BRIR with an IR particularly different in terms of frequency response and reverb time, the weighting coefficient of the cross-synthesis process may be varied

(typically, it would be diminished) in order to prevent the alteration of important spatial information within the BRIR.

Another important issue is related to the fact that for the distance simulation technique presented in Section 6.3, the BRIRs are divided into two different sets, an early reflections and a reverberant set. In this case, the cross-synthesis process has been applied only to the reverberant BRIRs set. An approximation was made in terms of the early reflections set, performing only for the early reflections BRIR a simple static equalization filtering process. A simple spectral multiplication which always applied 0.5 as the weighting coefficient for the filter signal and kept the phase information of the BRIR, i.e., similar settings to those listed in the previous section, was made between the early reflections BRIR and the correspondent early reflections components of the IR, using one single window for the whole process. No spectral envelope filter was applied to the early reflections components, given that the early reflections arrival time might vary depending on the simulated environment.

Further, it can be noted that no parameters were established within this simulation for the temporal characteristics of the early reflection, for example, the delay between the direct path signal and the arrival of the first reflection or the length of the whole early reflections "zone"; these vary according to the acoustic characteristics of the environment simulated. In this case, too, an approximation was made, incorporating those temporal factors considered less relevant than those of the spectral envelope. The timing of the three different IR components was kept the same as that of the BRIRs measured taking into consideration the possible alterations described in Section 6.3.2.

Further tests on the filtering of the early reflections spectral envelope and temporal characteristics are required in future studies, as mentioned in Chapter 10.

6.5 Brief summary

In the present chapter two intimately related innovative techniques have been presented and analysed. To simulate the distance of a virtual sound source, the anechoic source signal to be spatialized was processed in parallel with performing a convolution with three different HRIRs and BRIRs sets, i.e., the direct path signal, early reflections and the reverberant, corresponding to the specific angles of azimuth and elevation of the position to be simulated. For the purposes of simplicity, only 0° of elevation was

considered). The convolution with the direct HRIR, created by isolating the direct component of the pseudo-anechoic HRIR, was made through performing a linear interpolation between the HRIR measured at different distances. The two BRIR sets were processed through the cross-synthesis algorithm in order to simulate different required acoustic environmental characteristics, then a parallel convolution was performed with the source signal to be spatialized.

The three generated spatialized signals were then combined, weighting the multiplication coefficient of each according to the distance to be simulated, and altering the predelay of the early reflections. Finally, the output signal was then processed through a gain reduction line and a low-pass variable equalization filter, in order to simulate the frequency-dependent and -independent components of the air absorption.

This chapter has introduced techniques generated through performing approximations and generalizations, as shown, for example, in Section 6.4.2. They need to be seen not from the precisely mathematical and DSP perspective, but from the perceptual point of view. As was stated in the introduction to this research (*see* Chapter 0), convolution reverbs are generally considered as being more suitable in terms of the quality of spatialization as compared with other digital reverb algorithms; however, there does exist a weakness. It lies in the lack of flexibility in terms of the acoustic characteristics of the simulated environment. The techniques presented here were elaborated while attempts were made to keep the same flexibility of non-convolution reverb algorithms within a BRIR-based binaural acoustic environmental simulation. The results of the perceptual tests performed on them seem to confirm that this method is actually successful, from the point of view of the quality of spatialization.

It must again be emphasised that extensive studies into the binaural perception of distance, as well as into the binaural perception of acoustic environmental characteristics, have not yet been carried out; within this research, numerous materials have been measured and generated, thus further development and testing stages are imperative if knowledge of these kinds of perceptual and DSP processes is to continue to be expanded (*see* Chapter 10).

Chapter 7

7. The Binaural Spatialization Tool

In the context of the simulation of 3D soundfields over headphones, the previous chapters have presented and closely analysed innovative techniques for the simulation both of the positioning angle and of the distance of a given virtual sound source, as well as for acoustic environmental simulation. The majority of this was described in Chapters 4 and 6.

The outcomes of this research project should be made available to third-party users, in particular to musicians and composers. Within this chapter the implementation of the techniques previously introduced will be discussed. The first Section examines the need for two different software frameworks, one for lower quality real-time binaural spatialization, and one for higher quality offline processing. The two implementations differ in various elements and functions, and also in the fact that one can work in real-time and the other cannot.

Section 7.2 describes the Ambisonic approach for binaural spatialization, one of the most important concepts related to the implementation of the two pieces of software; separate sections will each present a deeper analysis of the two versions.

7.1 Real-time and offline version

As described in Chapter 3, convolution is a relatively simple mathematical operation although from a computational perspective it is far from efficient. The number of basic operations (sum and shift, therefore sum, subtraction, multiplication and division) required by the CPU to perform the function is exponentially linked with the length of both the signal to be processed and the IR. Given an IR h and another k doubles in length, if for $x \otimes h$ the number of operations to be performed is z, for $x \otimes k$ the number of operations is not 2z, but z^2.

The techniques presented in Chapters 4 and 6 for the simulation of the angular positioning and distance of a virtual sound source within a 3D soundscape over headphones require the execution of different parallel convolutions with long-length IR. Three different convolution processes need to be performed in simulating a virtual sound

source, using HRIRs and BRIRs of various length, included among 256 samples (the pseudo-anechoic HRIR) and up to ~100,000 samples (the reverberant BRIR).

A simulation performed according to those techniques would therefore require a significant number of calculations. It is true that the convolution can also be performed in the frequency domain instead of in the time domain, as is shown in Section 7.3, below. The complex sample-to-sample multiplication scheme in the time domain becomes a simple multiplication between FFT windowed spectra in the frequency domain. Nevertheless, in this case, too, the number of operations demanded in achieving real-time spatialization would be considerably greater than the possibilities offered by today's standard PC.

The real-time features of the algorithm to be implemented must be taken into consideration. As was stated in the introduction to this Ph.D. thesis, the software is assumed to be used mainly by composers and musicians to perform easily both binaural spatialization and conversions between multichannel formats and binaural. To achieve these, direct monitoring of the processed materials would be required; that is, the opportunity of listening to that which is being processed needs to be provided to the user. Buffers must be created for the material to be processed; the signals then need simultaneously to be played back so that any delay generated by the processing itself is minimized. This requirement renders the implementation of the presented techniques far more complex, if not impossible, for a real-time processing engine based on a standard PC. Of course, in future, with more powerful PCs, the situation will most probably improve. The only way to reduce the calculation load and thus allow a real-time processing with standard CPUs would be to approximate the processing. This could, for example, be applied to the reverb simulations, drastically reducing the performance in terms of the quality of spatialization. Chapter 2 listed and analysed various real-time binaural spatializators; approximations in terms of spatialization processing have always been linked to a substantial decrease in the quality of spatialization.

For this reason, the idea of splitting the implemented algorithms into two separate parts has been most carefully considered. A primary real-time binaural engine with simplified binaural processing, thus a lower quality of spatialization, is implemented in order to allow to the composer and/or the musician to monitor the processed signals during the compositional stage. A higher quality binaural engine is then programmed, implementing all of the spatialization techniques presented within this thesis, for an offline

processing to be performed once the signals have been edited, spatialized, and mixed into an intermediate (in this case, a Second Order Ambisonic signal) audio format.

In order to clarify the workflow to be followed when using both of the binaural processors, a simple example can be given of a compositional process:

- The composer uses the real-time binaural engine to create individual static or moving virtual sound sources within a 3D soundscape. It will be possible for him/her to monitor in real time the generated soundscape both in binaural (using simplified room simulation rendering) and multichannel (deciding between different loudspeaker set-ups) modes.
- Once all of the virtual sources have been created and mixed, within the real-time binaural engine the user is given the possibility to bounce down the audio material in an Ambisonic format (typically, of the Second Order).
- The nine-channel Ambisonic files may then be imported into the offline binaural engine, and converted to binaural through a more accurate processing, with full distance and environmental acoustic simulation functions. Here, the user is asked to choose which environment and which loudspeaker set-up are to be simulated, and to select the amount of reverb to be added to the signal. The final binaural file has therefore been created.

It should be borne in mind that the whole application, both offline and real-time, has been implemented and compiled for Intel Mac (Universal Binary) platforms. In the future it will be possible to transfer some or all of the tools to Windows-based systems.

7.2 The Ambisonic approach to binaural spatialization

As the previous chapters have stated, in order binaurally to spatialize a signal coming from a virtual sound source it is necessary to perform a convolution between the signal itself and the HRIR corresponding to the position to be simulated. For the spatialization of multiple virtual sound sources, one convolution for each of these situatiuons needs to be performed; for the simulation of the movements of a virtual source, the trajectory needs to be sampled, then HRIR convolutions to be performed for each step. An interpolation among all positions is thus achieved. In actual fact, the processing required in simulating moving sound sources is far more complex; for further information, see Matsumoto (2003).

Nevertheless, different approaches may be taken in order to perform these simulations, one of which will be described in the following lines. Instead of creating directly a binaural soundfield for each source then combining the different spatialized signals, an Ambisonic soundfield (see Section 2.6) encoding all of the different sources, including moving ones, can be created in First or Second Order format, and converted into binaural. This approach, already presented in McKeag (1996) and Noisternig (2003a; 2003b) allows higher efficiency for complex soundfields, because adding one virtual source results in a very small increase in the number of calculations to be performed. Also, higher flexibility is achieved thanks to the fact that the encoding process of the virtual sound source is separated from the binaural conversion process. To this should be added the fact that the simulation of moving sound sources may easily be carried out in the Ambisonic domain, converting the results into binaural without the need for complex HRIR interpolation operations.

To explain this particular approach, Figure 1 shows the workflow of a simple First or Second Order implementation:

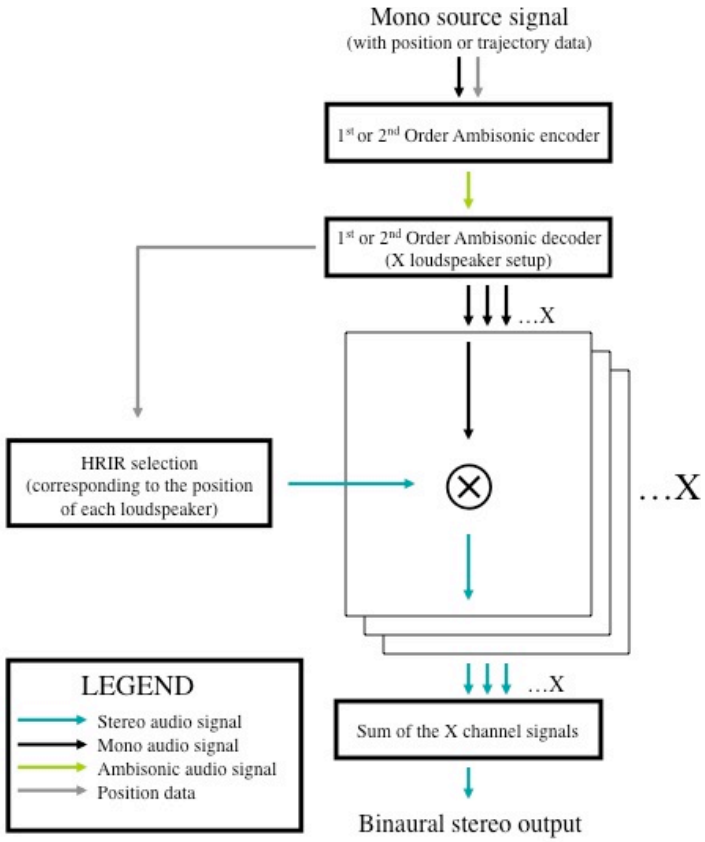

Figure 1. A schematic view of the architecture of a binaural spatializator based on the Ambisonic approach.

As in the diagram, the signal of a virtual sound source is encoded, following static or dynamic (trajectory) positioning data, into a First or Second Order Ambisonic format (four or nine channels, respectively). The Ambisonic soundfield is then decoded into any typology of loudspeaker set-up, be it bi- or three-dimensional, with speakers placed at the corners or angles of the array, for example, in a square, cube, dodecahedron or hexahedron. Each of the loudspeaker signals is convolved with the HRIR corresponding to its position before all of the binaurally spatialized loudspeaker signals are combined into a single stereo signal, which is the binaural encoding of the simulated soundfield.

This approach, as has been stated, allows both a more efficient binaural process for highly complex soundfields, and a more flexible management of the audio files between the two versions of the binaural software implemented within this research; in fact, the composer and/or musician is able to create any kind of soundscape with the real-time binaural engine, monitoring it with a simple Ambisonic-to-binaural conversion, and encoding it into a Second Order Ambisonic audio file, i.e., of nine channels. The Second Order Ambisonic audio files are then imported into the offline binaural engine and converted into binaural, performing a higher quality encoding, with distance and environmental simulation, and outputting a high quality binaural stereo signal.

A precise description of the two different engines, with diagrams, features and functions, is given in the following sections.

7.3 Implementation of the offline binaural tool

Here, information is given on the implementation of the offline binaural engine. Starting by listing the different modules and functions of the software, an overview follows of how these have been implemented, of First and Second Order Ambisonic decoding equations, and of the BRIR characterization process.

7.3.1 The different modules and functions

Different modules have been created, each corresponding to a particular version of the binaural spatialization software. A list of these follows:

- **Source Positioning Binaural.** Binaural spatialization of one single source, with distance and environmental simulation: one monophonic channel input and one stereophonic binaural output. The module needs to be input with the positioning data (azimuth, elevation and distance) and with the environmental data (in this first ver-

sion, only three environments are available: small, medium and large), as well as with the filename and path of the mono audiofile to be spatialized and of the output file. A schematic view of this module is found in Figure 2:

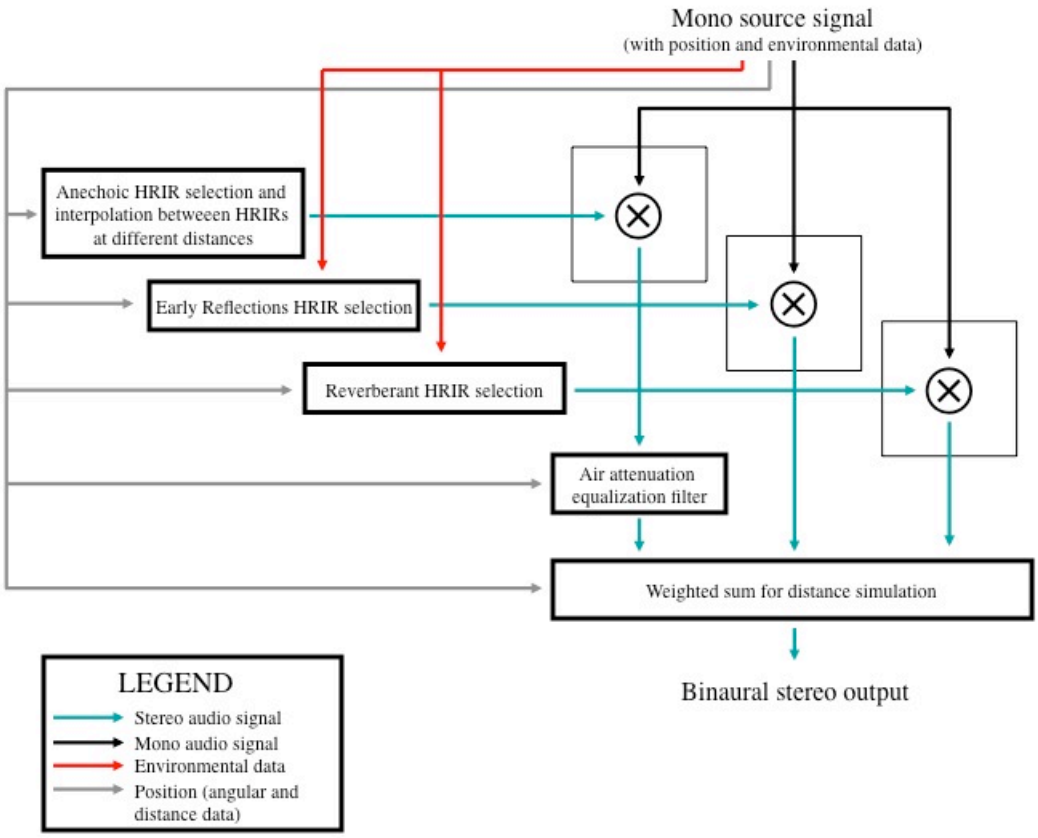

Figure 2. A schematic view of the architecture of the Source Positioning Binaural module

- **Multichannel-to-binaural conversion.** Conversion from any multichannel signal (in the form of multiple monophonic audiofiles each corresponding to one loudspeaker) to binaural, with the distance specified for each channel, and environmental simulation results in *x* monophonic-channel input and one stereophonic binaural output. This module is basically a combination of Source Positioning Binaural modules, one for each channel (with a final weighted sum for each of them), in order to convert to binaural any multichannel signal in the form of monophonic audiofiles, each corresponding to one loudspeaker. The module needs to be input with the positioning data (azimuth, elevation, and distance) for each loudspeaker of the array, followed by the filename and path of the corresponding mono audiofile, and with

the environmental data (in this first version, only three environments are available: small, medium and large), as well as with the filename and path of the output file. A schematic view of this module is found in Figure 3:

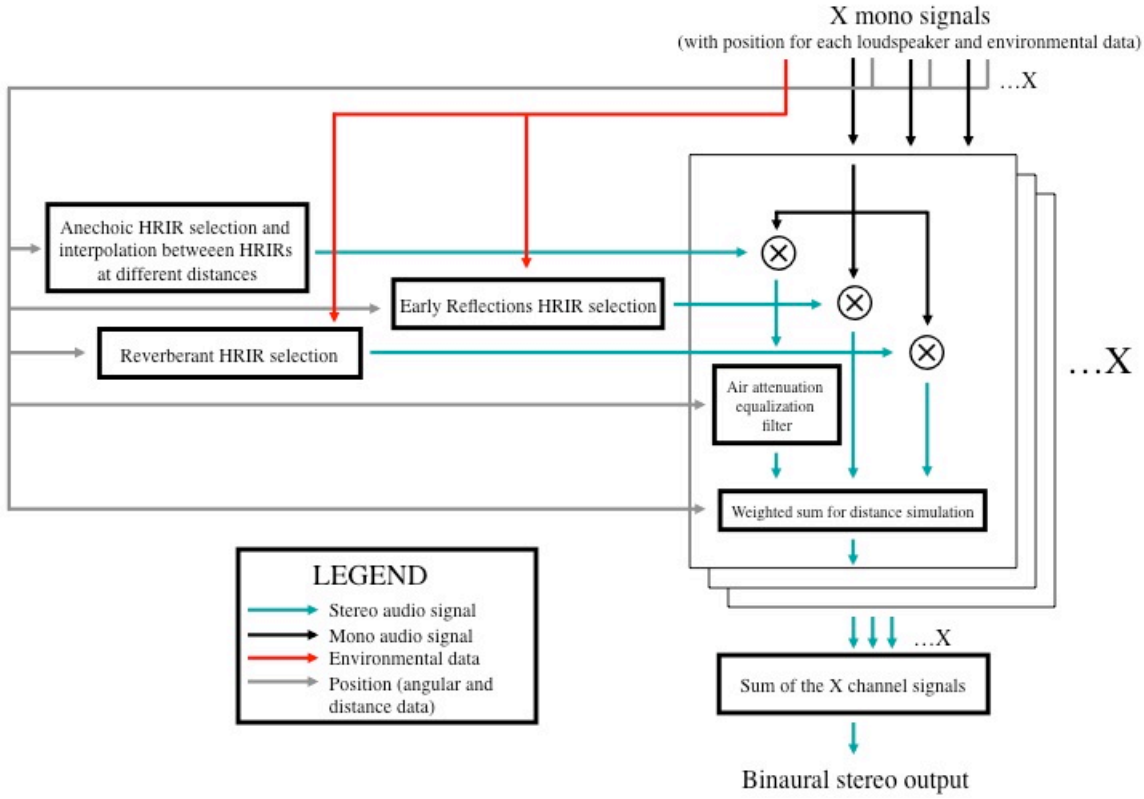

Figure 3. A schematic view of the architecture of the Multichannel TO Binaural module.

- **First and Second Order Ambisonic-to-binaural conversion.** Conversion from a First or Second OrderAmbisonic signal,in the form of multiple monophonic audiofiles, each corresponding to one Ambisonic channel, to a binaural, with distance (for each channel) and environmental simulation require four (First Order) or nine (Second Order) monophonic-channel input and one stereophonic binaural output. This module is very similar to the previous one (multichannel-to-binaural); here, however, an Ambisonic decoder is implemented before the multichannel conversion section. For more information about the Ambisonic decoding implementation, *see* Section 7.3.3. The four- or nine-channel Ambisonic signals are first decoded into a loudspeaker array, of which three choices are available: Octagon 2D, Cube 3D and Dodecahedron 3D. Subsequently, each channel (corresponding to

a loudspeaker) is converted into binaural, and finally combined in the binaural stereo output file. The module needs information to be input; this consists of the Ambisonic Order (First or Second) and the loudspeaker array set-up (Octagon 2D, Cube 3D or Dodecahedron 3D, above), with the filenames and paths of all of the Ambisonic channels given in the standard Ambisonic Order, therefore W-X-Y-Z for the First Order, plus R-S-T-U-V for the Second Order, and with the environmental data. In this first version, only three environments are available: small, medium and large. The filename and path of the output file must be input. A schematic view of this module is found in Figure 4:

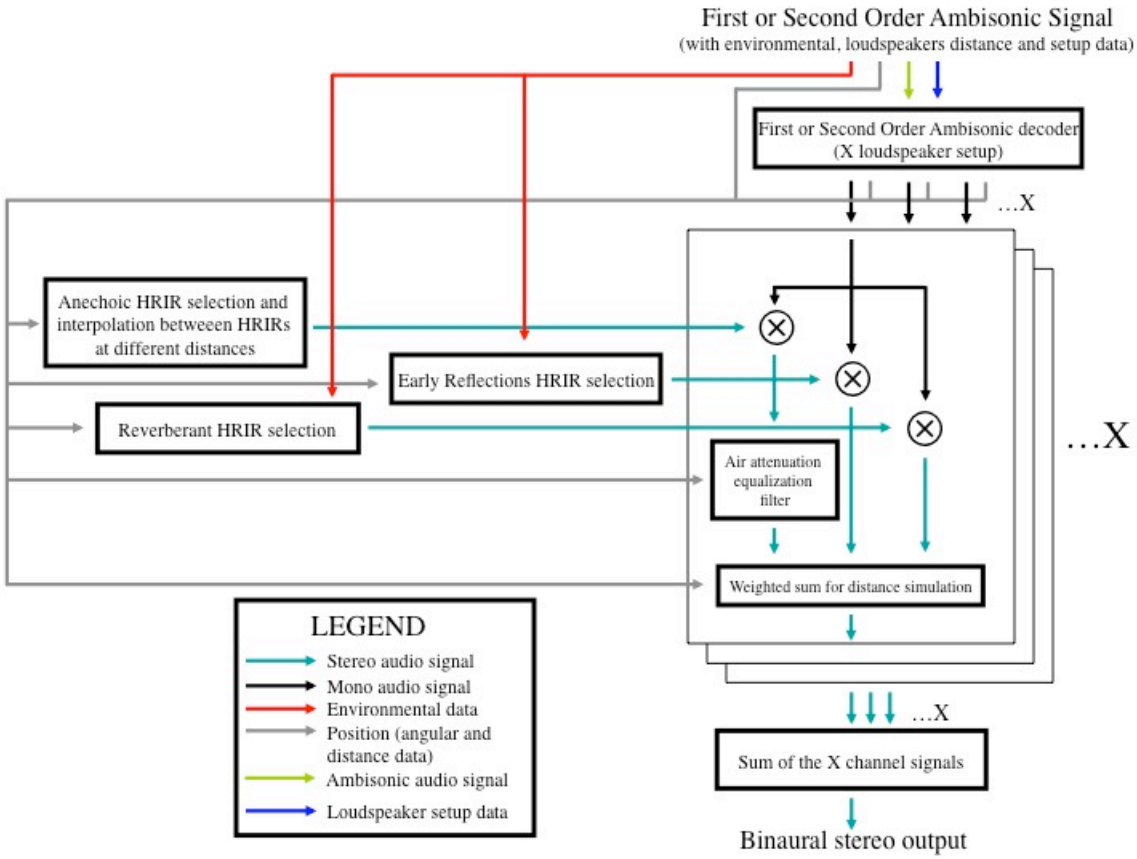

Figure 4. A schematic view of the architecture of the Ambisonic-to-binaural module

As an additional function, each of the three modules also outputs the three stereophonic signals corresponding to the binaural spatialization performed only with direct path HRIR, with the early reflection BRIR and with the reverberant BRIR. The user will thus

be able either to use the combined stereophonic binaural output, or him/herself to combine the three binaurally spatialized components.

7.3.2 Implementation of the different functions

The whole software package was programmed in C++, using the Apple XCode[1] programming environment and the gcc compiler (*see* Wall, 2004). For the audio file reading and writing, the soundfile++[2] library was applied; note that the use of this library limits the format for the input and output audiofiles to Microsoft Wave. The FFT for the frequency domain convolution engine was performed using the fft2[3] C++ function.

A C++ convolution function named convolve was implemented and employed within all of the software modules. This function performs the convolution in the frequency domain between a signal and an IR through multiplying the spectrum of the signal with the one of the IR; it is achieved through using an FFT window with the dimensions of the length of the selected IR for both HRIR and BRIR. The convolve function receives the pointers to the mono audio signal to be spatialized and to the stereo HRIR or BRIR corresponding with the position of the virtual sound source to be simulated. It returns the pointer to the stereo audio signal generated after the convolution process.

Another C++ function was implemented for distance simulation: distSim mixes the signals spatialized with the different HRIR and BRIR components to different weights, dependent on the distance to be simulated, and modifies the pre-delay of the early reflection components. Further, an option is given to the user manually to enter the pre-delay value, independently of the distance to be simulated, which filters the direct path signal because the filter is realized in the frequency domain, in order to simulate different air absorption coefficients. Further information on the principles of binaural distance simulation was provided in Chapters 5 and 6. The distSim function receives the pointers to the three stereophonic signals corresponding to the audiofile spatialized with the three HRIR and BRIR components, together with the data on the distance to be simulated, between a minimum of one metre and a maximum of twenty metres. It then

1 *See* http://developer.apple.com/mac/

2 *See* http://soundfile.sapp.org/

3 Written by Dale Carstensen, the Antares project, Los Alamos National Laboratory, 16 March 1981 for Unix version 6.

returns the pointer to the stereo audio signal generated combining and filtering the different input signals.

A third C++ function has been created for the First and Second Order Ambisonic decoding, named ambiDec. This function receives the pointers to the Ambisonic signals, the Ambisonic Order data, that is, whether they are of First or Second Order, and the decoding loudspeaker array set-up, consisting of Octagon 2D, Cube 3D or Dodecahedron 3D, as described in Section 7.3.3. It then returns the decoded loudspeaker signals, which are then individually converted to binaural. The decoding equations are also given in Appendix B.

The three software modules use therefore these three C++ functions in order to perform the various binaural processes: the input for the different parameters including audiofile names and paths, position coordinates, and Ambisonic Order is given through a command line interface, calling first the module name, then replying to the different questions automatically asked by the module, such as the name and path for each of the audiofiles, and the spherical coordinates for the source to be simulated.

The whole package is compiled for Apple Mac with Intel Core Duo processors, and uses the Terminal for calling the different modules and for inputting the data for the processing.

7.3.3 First and Second Order Ambisonic decoding equations

As was outlined in the previous section, within the implementation of the First and Second Order Ambisonic-to-binaural conversion module an Ambisonic decoding function was created to convert the Ambisonic channels into discrete loudspeakers channels. Different decoding equations, according to the Ambisonic Order and to the loudspeaker set-up, were used in order to calculate the weighting coefficients for the transformation matrix between the Ambisonic and the loudspeaker channels. The Second Order format used is the Furse-Malhalm Higher Order Format (*see* Malham, 1999; 2003), which provides a full 3D Second Order Ambisonic implementation, and the coefficients are the same as that used by the MN C++ Audio Library (MNLib) programmed by Richard W. E. Furse[4] (*see* Appendix B).

4 *See* http://www.muse.demon.co.uk/ref/speakers.html

The loudspeaker positions (azimuth and elevation) for the three loudspeaker configurations, plus a fourth used only for the real-time version, and for which the apex positions have been stretched in order to correspond with the measured HRIRs, are given in Table 1.

Loudspeaker	Octagon 2D	Cube 3D	Dodecahedron 3D	Icosahedron 3D (only for the real-time version)
1	22.5° 0°	45° -35.25°	0° 90°	90° 60°
2	67.5° 0°	315° -35.25°	0° -90°	-90° 60°
3	112.5° 0°	225° -35.25°	40° 26.55°	90° -45°
4	157.5° 0°	135° -35.25°	216° -26.55°	-90° -45°
5	202.5° 0°	45° 35.25°	320° 26.55°	60° 0°
6	247.5° 0°	315° 35.25°	144° -26.55°	-60° 0°
7	292.5° 0°	225° 35.25°	108° 26.55°	120° 0°
8	337.5° 0°	135° 35.25°	288° -26.55°	-120° 0°
9	-	-	252° 26.55°	0° 30°
0	-	-	72° -26.55°	0° -30°
11	-	-	180° 26.55°	180° 30°
12	-	-	0° -26.55°	180° -30°

Table 1. Loudspeaker positions for the three loudspeaker configurations, plus a fourth one used only for the real-time version.

7.3.4 BRIR characterization

In reference to that which was outlined in Chapter 6 (Section 6.4), three different sets of BRIRs have been created using the cross-synthesis process, starting from the Trinity Chapel measured BRIRs. They have been in fact been filtered (for more information on the cross-synthesis process, *see* Section 6.4.1) with three IRs calculated using an acoustical model done with the CATTAcoustic[5] software.

The characteristics of the simulated environments in terms of RT60 for different frequency bands are found in Table 2.

5 *See* http:///www.catt.se

Name	X	Y	Z	125Hz	250Hz	500Hz	1kHz	2kHz	4kHz
Small	3	3	2.3	0.44	0.38	0.37	0.35	0.38	0.38
Medium	5	5	2.5	0.95	0.99	1.18	1.15	1.12	0.88
Large	10	10	3.5	1.58	1.60	1.94	1.88	1.83	1.35

Table 2. Size in metres (for the three axes) and RT60 in seconds for six different bands of the three simulated environments, used for the cross-synthesis filter for the characterization of the BRIRs.

It is important to note that the multiplier used for the cross-synthesis process for the filtering signal (in this case, the 3 CATTAcoustic IRs) is 0.5; therefore, the resulting RT60 for the three characterized, that is, processed, BRIRs are slightly longer than those listed in Table 2.

7.3.5 Problems and final organization of the tool

It is worth citing here some of the problems encountered during the implementation of the different functions and modules:

- Problems have been found at the beginning of the programming process for the choice of the sound input library. The soundfile++ library was considered the most simple and flexible for this simple audio input and output need, and was therefore included within the code. The same issue was found for the FFT function, and for the same reasons the fft2 function was used.
- The first implementation of the convolution function was performed in the time domain, resulting in a processing that was far from being efficient, most of all for the room simulation module (with BRIRs longer than 500 ms). The function was therefore re-implemented in the frequency domain, resulting in a massive shortening of the processing time.
- The first version of the software was implemented in a unique function, resulting in a more complex management of the different code parts, particularly regarding the re-use of the same functions for other modules. Three separate functions, as described in Section 7.3.2, were therefore created, allowing the different modules to call the required functions without needing to copy and paste code parts within the main function.

The offline binaural spatialization tool was finally organized as three-module software. According to the function requested (*see* Sections 7.3.1 and 7.3.2), the user simply needs to call from the Terminal window the corresponding software module, to follow the instructions appearing on the screen, and to wait for the process to be completed.

7.4 Implementation of the real-time binaural tool

As was pointed out in Section 7.1, the offline version of the binaural tool has no real-time monitoring functions. The audio is input into the different modules, processed, and output, with no possibility of monitoring the result other than by listening to the output files once the processing is complete.

This would probably create no problem when a conversion between already mixed multitrack signals and binaural is to be performed; in this specific case, the audio would already have been mixed using multichannel loudspeaker arrays, and would merely need to be converted to binaural. Of course, adjustment to parameters might be required, and this can easily be executed simply through listening to the output files and re-calibrating the algorithm parameters.

On the other hand, when audio needs to be processed directly into the binaural domain through using, for example, the Source Positioning Binaural Module to create different virtual sound sources, real-time monitoring becomes an essential feature. Additionally, there is also the fact that in the Source Positioning Binaural Module no moving sound source functions are implemented, and therefore the simulated virtual source can only be static. The user would therefore need to create, using any multichannel audio editing and mixing software, a multichannel audio mix with static and/or moving virtual sound sources, allowing him/her to pan, mix, and monitor the sources in real time and, once the mix is ready, to convert it into binaural, this time using the Multichannel-to-binaural module. In order properly to accomplish this, the user will need to have a well-configured 2D or 3D surround sound system, with the number of loudspeakers required by the chosen set-up and with a proper listening environment. The placing of loudspeakers for surround sound set-ups will be discussed in Chapter 9. The various features required are available only in specific structures, such as in universities or recording studios, and are most probably not accessible to every individual.

For these reasons, a binaural real-time software module was created using MaxMSP.[6] The main aim of the implemented platform is to assist the user in the creation of 2D and 3D multichannel audio soundscapes offering Ambisonic encoding functions and binaural monitoring.

7.4.1 MaxMSP

This section provides a very brief introduction to the software used for the development of the real-time binaural platform. MaxMSP is a visual programming language for music and multimedia developed and maintained by Cycling 74 (San Francisco, USA). It is a highly modular programming environment, which takes advantages of a large community of programmers that has developed, and is still developing, a large number of objects, functions and library for signal processing within the main application.
The choice of this software environment for this specific task may easily be justified by the following factors:

- High flexibility of the programming environment, including the possibility easily to create GUIs (Graphical User Interfaces) to interfacie the software core with the final user
- Cross-platform usage of the implemented software, which can easily be used in PC and Mac operative systems without any particular conversion
- A high number of available libraries and objects for different typologies of signal processing, including Ambisonic encoders, decoders, and convolution functions.

Even though in 2008 a new version of MaxMSP became available (MaxMSP version 5), version 4.6 was used for the implementation of the real-time binaural platform. The choice can be justified through considering the differences between the two versions: a completely new graphic interface, different frameworks for the programming of the various objects, and different management of inter-object links are some of the factors. Many already existing objects ("externals") were used in the development and implementation of this application (for example, for the Ambisonics encoding and decoding processes); however, for many of these compatibility with MaxMSP version 5 was not supported It was, therefore, considered more practical to implement the whole platform using MaxMSP version 4.6 (specifically, version 4.6.3), allowing for the possibility to

6 *See* http://www.cycling74.com

"transfer" the whole work into MaxMSP 5 when all libraries used are available for that version.

7.4.2 The different functions

In order to allow the user to create 2D and 3D soundfields through the Ambisonic encoding of imported soundfiles, monitoring the process in real-time, the following functions were implemented within the MaxMSP platform:

- Second Order Ambisonic encoder for multiple input channels (signals are played back through multiple mono audiofile players), with the possibility of inputting fixed polar coordinates or 2D trajectories.
- Simplified (Second Order for the direct signal, and First Order for the early reflections and reverb) Ambisonic-to-binaural conversion, based on a virtual loudspeaker array, for monitoring purposes using two different loudspeaker set-ups: a 3D icosahedron (*see* Section 7.3.3) for the direct signal, and a 2D square for the early reflections and reverb.
- Flexible management of the weighted sum between the signals spatialized with the different HRIR and BRIR components, including simulation of pre-delays for both the early reflection and reverb signals.
- 9 channels audio recorder for exporting the Ambisonics soundfields created to the offline conversion modules.

A schematic view of the architecture of the real-time Ambisonic-to-binaural platform is given in Figure 5.

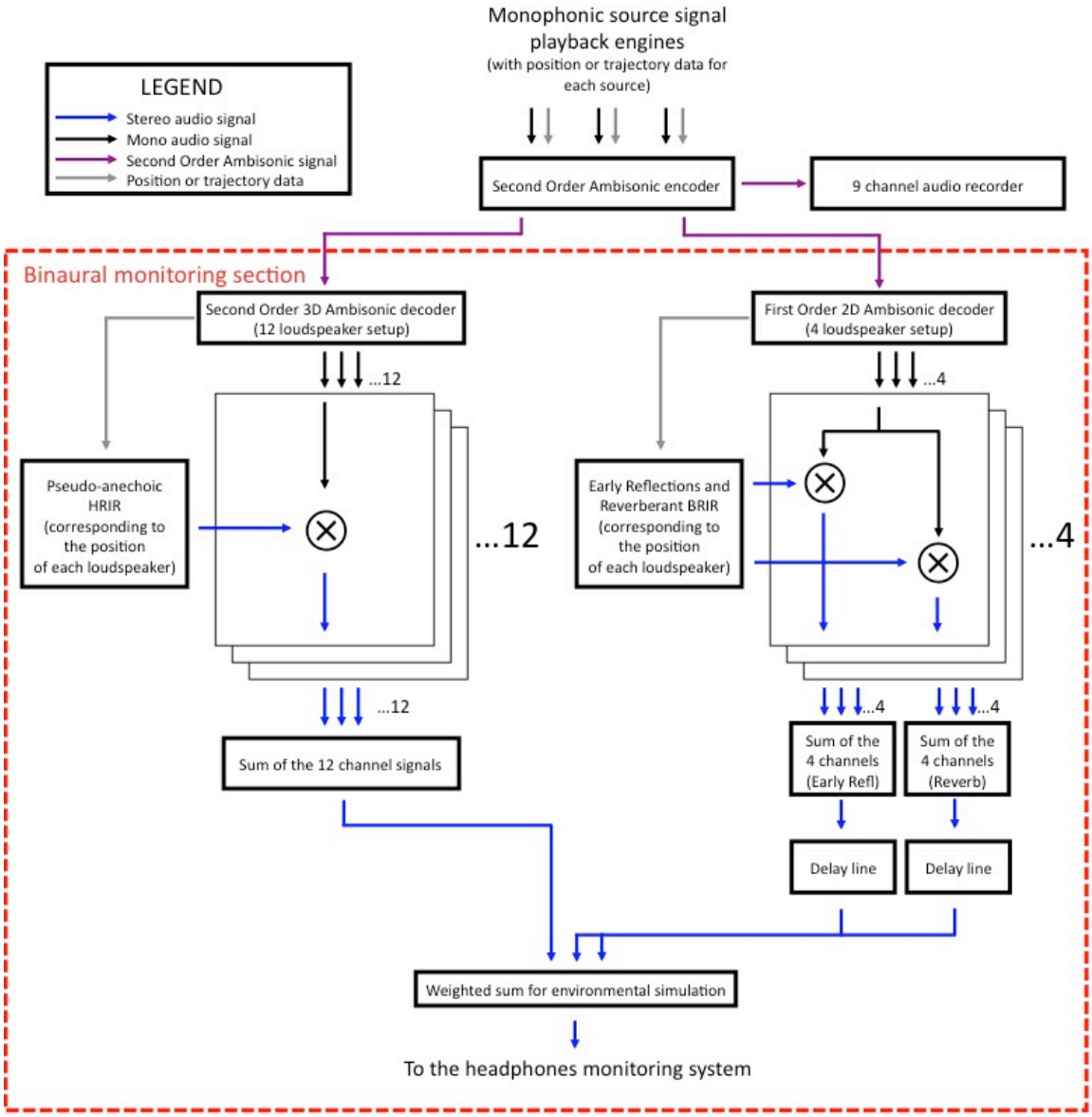

Figure 5. A schematic view of the architecture of the realtime MaxMSP Ambisonic-to-Binaural platform.

7.4.3 Implementation of the different functions

The different functions outlined in Section 7.4.2 were implemented using MaxMSP objects and libraries freely available in the Cycling74 installation package and online. The playback of the mono soundfiles was implemented using the sfplay~ object, while for the drawing of the trajectory data the Trajectory object, from J. B. Thiebaut,[7] was

7 Of the Centre de Recherche Informatique de Création Musicale, Paris, France.

used. The Ambisonic encoding was performed using the ICST ambiencode~ object [8], and the decoding using the Graham Wakefield Ambisonics for MaxMSP library[9]. The nine-channel recorder was implemented using a multichannel version of the sfrecord~ object.

For the direct path components of the Ambisonic-to-binaural conversion, twenty-four instances (twelve of them stereo) of the buffir~ object were applied in convolving the loudspeaker channels coming from the Ambisonic decoding with the HRIRs (in this case, only 128 sample HRIRs were used, in order to optimize the real-time processing) corresponding with the actual loudspeakers' positions. The buffir~ object allows convolution with a maximum 256-sample IRs; therefore, for convolution with BRIRs for the early reflection and reverb simulation, another object was employed: the TConvolutionUB~ v. 0.1 by Thomas Resch[10] (which allowed the use of >44,100 samples IRs). In this specific case, eight instances (of which four stereo) of the TConvolutionUB~ object were used for both the simulation of early reflections and reverb.

For the simulation of the early reflections and reverb, the user can freely select any or all of the three sets of characterized BRIRs (small, medium and large, *see* Section 7.3.4).

The coefficients for the weighted sum of the signals spatialized with the three HRIR and BRIR components, as well as the values of the delays (implemented using the delay~ object) applied to the early reflections and to the reverb signals, can be manually set up by the user. Three different presets are nevertheless present for each of the room simulations.

A GUI was then created in order to render all functions and parameters easily accessible and selectable.

7.4.4 Final organization of the tool

The whole software platform was finally organized into a single folder, to be imported and copied in any part of any hard drive connected with the computer of the user. To use the real-time binaural software platform, the user will need to install a full or runtime (freely available in the Cycling74 internet site) version of MaxMSP.

8 *See* http://www.icst.net

9 *See* http://www.grahamwakefield.net

10 *See* http://www.zippernoise.net/download/tconvolution~MaxMSP.zip

All of the HRIR and MaxMSP software objects essential for the functioning of the platform are already present within the main application folder. The user will need only to open the main MaxMSP file and run the application.

7.5 Low Frequency Compensation

After some informal perceptual evaluations of the developed applications, it has been found out that the whole binaural spatialization process creates a loss in terms of frequencies below 80-100 Hz, with the results of having spatialized signals with very weak low frequencies content. These problems are probably due to the calibration problems outlined in Section 4.3.2. In order to compensate this frequency loss, a *Low frequency compensation* function has been added to all the developed applications (both offline and real-time). This function works in parallel with the whole spatialization system, and is inputted with the "dry" signal, just before this being sent to the spatialization algorithm. This is then filtered with a variable low-pass frequency filter and summed to the spatialized signal, just before this being outputted. Varying the level of the filtered signal to be summed with the spatialized one, allows compensating the low frequency loss.

7.6 Brief summary

In this chapter, the organization and implementation of the real-time and offline binaural processing software have been analysed and described. The reason why two different items of software were created can be justified by two significant factors: firstly, the requirement in terms of CPU calculations of binaural processing with distance and environmental simulation is too large to allow a real-time version of the different modules, and therefore a real-time lighter implementation of the algorithm is necessary for monitoring purposes. Secondly, the offline binaural modules do not possess moving sound source functions; these thus need to be created in the Ambisonic domain, allowing real-time monitoring, using again a lighter implementation of the algorithm, and then converting the signals to binaural using the more complex offline modules.

Chapter 8

8. Subjective Perceptual Tests

In the previous chapters, the mechanisms for the binaural perception of the angular position and of the distance (Chapters 3 and 5) of sound sources were introduced, as were the simulation of localization and distance cues within a binaural spatialization algorithm (Chapters 4 and 6). In Chapter 7, a whole binaural tool was implemented for allowing a composer-musician – and, more generally, any individual – to use the techniques developed for the simulation of three-dimensional soundfields over headphones. In particular, two techniques were outlined to represent the main innovations of this research: the simulation of the distance of a virtual sound source and, closely linked, the simulation of the acoustic characteristics of the environment, therefore the simulation of the early reflections and of the reverb.

In the present chapter, a description will be given of the development, implementation and analysis of the results of two subjective perceptual tests performed during and at the end of the research in order to verify the effectiveness of the techniques developed. A "work-in-progress" test was carried out in the first months of 2007 in order to test the efficacy of the first implementation of the distance simulation technique, and a second, more thorough, test at the end of the research and development stage comparing the benefits of the binaural environmental simulation technique with other 3D, 2D and 1D environmental simulation techniques over headphones. Therefore, in the first two Sections of the current Chapter the two tests are described, and a brief analysis of the results is provided; next, an overview of other possible perceptual tests on the applications developed is reported, and conclusions drawn from the whole subjective perceptual test stage are summarised.

All of the analysis of the tests and the results charts has been computed using Matlab (for the data collection and analysis, and for the diagrams) and Microsoft Excel (for the tables).

8.1 Distance simulation test

As has already been explained in the introduction to this Chapter, this first test was carried out in the first months of 2007, just after the earliest development and imple-

mentation of the first version of the distance simulation algorithm. After this test, many modifications were made on the distance simulation technique, as are found in Chapters 6 and 7. It is nevertheless true that the test described here was essential to examining and evaluating the main ideas behind the technique. The first version of the binaural reverb and distance simulator implemented in MaxMSP (*see* Section 8.1.1, below) was not oriented towards a proper simulation of all the parameters involved in the estimation of the distance of a sound source; it was simply a pilot study to verify how the changing of the different parameters (the weighting of the different HRIR and BRIR components, air absorption, etc.) could cause a variation in the distance perceived.

In Chapter 6, an accurate description of the distance simulation technique, based on the splitting of the different HRIR (and BRIR) components, performing then a weighted mix between the signals spatialized using these, was carried out. The initial perceptual test was planned and performed in order to test the idea of varying the simulated distance, replicating individually the different distance cues, starting from an HRTF measured at one metre and then simulating larger distances.

It is important to underline that while for the second (and more important) test described in the current chapter (*see* Section 8.2) an extensive statistical analysis of the results has been carried out, the results for this "work-in-progress" test have been analysed only through observing the boxplots (*see* Charts 1, 2, 3 and 4).

Some of the results of this test have already been published by the author (Picinali, 2007a; 2007b).

In order to obtain significant results, the test has to be performed with a minimum of fifteen to twenty subjects.

8.1.1 Binaural reverb and distance simulation: first MaxMSP implementation

After outlining the basis of the binaural reverb and distance simulation technique (*see* Chapter 7), an initial implementation was created using MaxMSP (*see* Section 7.4.1). Within the application developed, spatialized signals, processed using different HRIR and BRIR components, were entered with different weights, then a simulation of the effect of the frequency-dependent and frequency-independent air absorption components was performed, each one relative to each distance to be simulated.

The spatialization of the signals was carried out offline (using Adobe Audition software with Aurora Plugins; *see* Section 4.2.4), and the following signals imported into MaxMSP:

- Original signal (non-spatialized, diotic stimulus)
- Anechoic signal (spatialized with the anechoic HRIR)
- Early Reflection signal (spatialized with the early reflections HRIR)
- Reverberant signal (spatialized with the reverberant BRIR).

In the first case, the early reflection signal was spatialized using the Early Reflections HRIRs (*see* Section 6.2.3), and the reverberant signal using the IOCT "covered" BRIRs (*see* Section 6.2.2). It should be noted that these are not the IRs used for the final version of the binaural tool (*see* Chapter 7); however, at the moment when the technique was first developed, these were the only available HRIRs and BRIRs. The Trinity Chapel BRIRs were measured only in the first months of 2009.

As is possible to see in Figure 1, within the MaxMSP interface the first slider corresponds to the simulation of the distance, and it controls the changes in the weights of the signals spatialized with the different HRIR and BRIR components, the cut-off frequency for the low-pass filter and the whole signal reduction. Starting from the original non-spatialized signal alone, i.e., the diotic stimulus, the anechoic HRTF components were added, then the early reflections and the reverberant ones. The gain of these last items increases even when the anechoic components are reduced. The following list shows the different multipliers of the differently spatialized signals changing with the distance, as well as the cut-off frequency. Between the different slider positions, the multipliers are interpolated linearly.

- Distance 0: only original signal, IHL
- Distance 200: anechoic signal 1, early reflections 0.2, reverberant signal 0.1 no filter
- Distance 400: anechoic signal 0.4, early reflections 0.6, reverberant signal 0.33 no filter
- Distance 600: anechoic signal 0.26, early reflections 0.4, reverberant signal 0.55, whole signal 0.833, low-pass 14333 Hz
- Distance 800: anechoic signal 0.13, early reflections 0.2, reverberant signal 0.77, whole signal 0.66, low-pass 8666 Hz
- Distance 1000: only reverberant signal, whole signal 0.5, low-pass 3028 Hz.

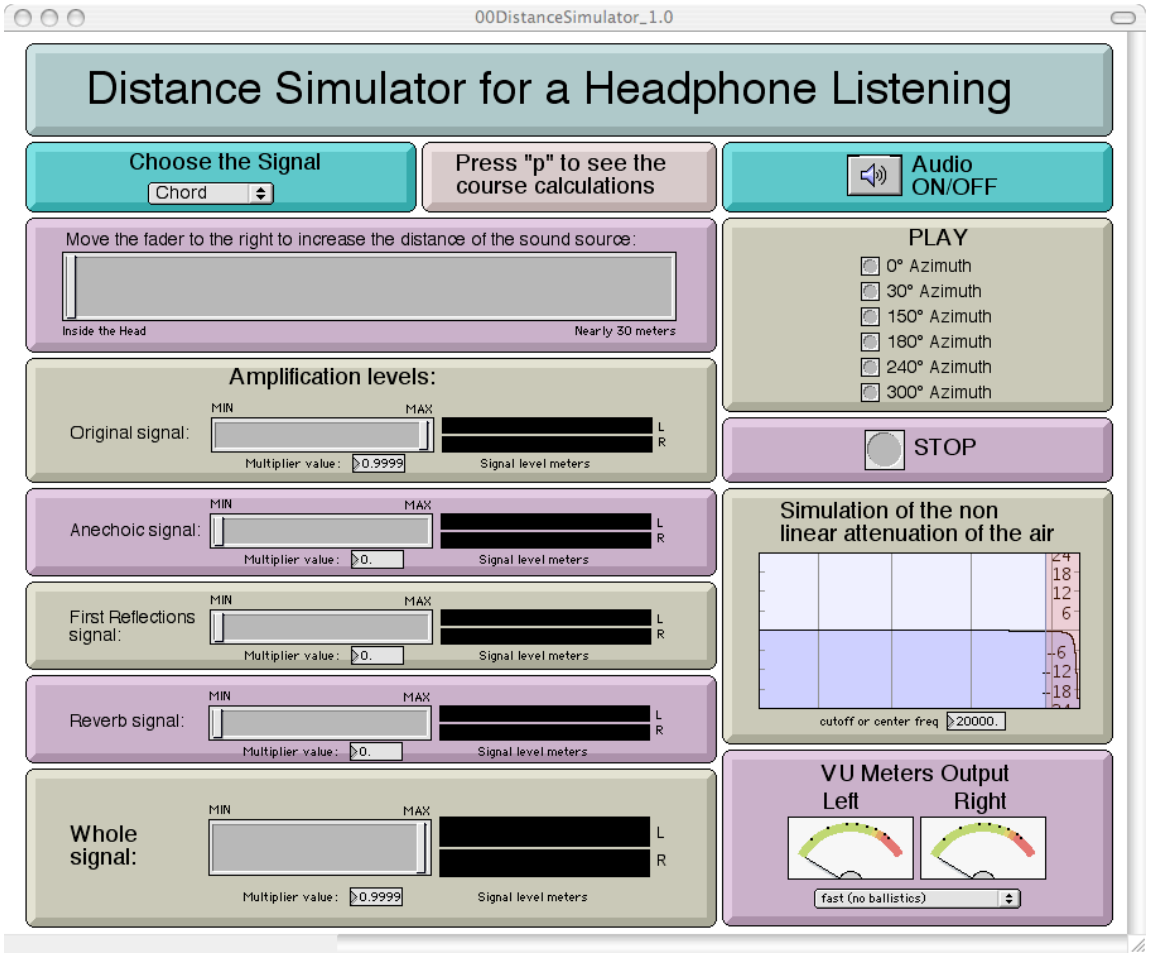

Figure 1. The Distance Simulator implemented in MaxMSP.

8.1.2 Objective

The objective of this first perceptual test is to validate and obtain feedback about the earliest implementation of the binaural reverb and distance simulation technique. Through simulating different distances or, even better, changing the levels at the differently spatialized components, and through reducing the cut-off frequency of the low-pass filter, the perception of the distance of the simulated sound source should change, and the reported perceptions of the individuals tested should reflect such changes.

It is important to emphasise that here, no accurate attempt at reproducing the distance cues was made; the test constituted simply a verification that the “direction” taken from the development of the binaural tool is the most appropriate. For this reason, results are

analysed through considering not the absolute evaluations of the estimated distance, but only the relative entities among the different samples.

8.1.3 Organization of the test

The organization of the test can be summarised in the following points:

- Twenty different listening samples, all with the same source signal, a female voice speaking in Spanish, were spatialized at different degrees of azimuth and at different distances. A language not understandable by the subjects tested (all native English speakers, none of whom has studied Spanish) was selected in order to help them focus on the location of the source rather than on the actual content of the signal. An elevation simulation was not performed because the goal of the test was to obtain feedback on the distance simulation, and thus the addition of the elevation parameter could have compromised the perception of the distance.
- Three different questions were asked to each subject:
  - Where do you perceive the location of the sound source in the horizontal plane? The answer can be given through a two-dimensional panel, which simulates the horizontal plane, with the head in the middle and a cursor that can be moved in any direction. The subject has to move the cursor into the position where s/he thinks the sound source is located on the horizontal plane. The cursor can be positioned even inside the head (*see* Figure 2). The information of the position of the cursor is saved using two numbers (the horizontal and vertical coordinates), from 0-0 (back-left) to 100-100 (front-right).
  - At which elevation do you perceive the position of the sound source? The answer can be given through a one-dimensional slider, which indicates the elevation of the sound source in the vertical (or median) plane. The slider has 100 steps, from 0 (“Under the Head”) to 50 (“Height of the Head”), thence to 100 (“Above the Head”). The subject has to move the cursor into the position where s/he thinks the sound source is located in the vertical (or median) plane. Even if in this case, as previously stated, no elevation simulation is performed, it was considered important to ask to the subject to estimate this parameter as it is known that through using non-individualized HRTFs a variation of the elevation

perception, even using always HRIRs measured on the horizontal plane, may occur (*see* Begault, 2001, and Møller, 1996)

- At which distance do you perceive the position of the sound source? The answer can be given through a one-dimensional slider, which indicates the distance of the sound source. The slider has 100 steps, from 0 ("Inside the Head") to 30 ("Outside the Head"), thence to 70 ("Inside the Same Room") and to 100 ("Really Far Away"). The subject has to move the cursor into the position that corresponds to his/her estimation of the distance of the sound source.

- The subject has to listen to each sample, to make the choices for the questions listed here, then to save them. S/he can listen to the same sample as many times as s/he wishes, but just to one sample at a time; once s/he has saved the choice, it is not possible to go back. This is to prevent the subject from making comparisons between the different samples; each sample has to be listened to on its own, and the choices have to be made regarding that sample only.
- At the end of the test, the listener is asked to save all the choices in a file labelled with his/her name and age. Each listener is also requested to enter briefly into a specific part of the platform the relationship of the listener with the musical world (such as "musician", "composer", or "only a listener").

8.1.4 Organization of the listening samples

Regarding the organization and the order of the samples to be played back to the subjects, a few rules were considered:

- Certain distance values need to be repeated in more than one listening sample, and these not in any sequence. They were selected because of their particular importance for different reasons (listed below), and it is essential to have at least two judgements about them in different parts of the test. For example, the first and the tenth samples are at 0 (distance slider) distance, in order to represent a sort of reference at the beginning and in the middle of the test. The other distance values are reported in the following list:
  - 0 and 200 (important in verifying the movement of sound sources from inside to outside of the head)

    - 400 (important in verifying the effect of the IHL effect and of the first reflections)
    - 700 (important in verifying the effect of the decrease of the direct-to-reverberant signal ratio)
    - 1000 (important in verifying the perceived distance of the reverberant signal).
- Three series of distance values have to be reported in a specific sequential order, only once; this is because of the importance in the perceived sensation regarding the gradual increase of the distance simulation. These sequences are reported in the following list:
    - From 0 to 150, in steps of 50 (in order to verify the movements of the sound source from inside to outside of the head)
    - From 200 to 350, in steps of 50 (in order to verify the effect of the increase of the early reflections components)
    - From 1000 to 700, in steps of 100 (in order to verify the effect of the decrease of the direct-to-reverberant signal ratio, and of the simulation of the frequency-independent components of the air absorption).
- The simulation of the spatialization on the horizontal plane (azimuth simulation) is not relevant for the goal of the test, although it is important to give different references at different azimuth positions. The azimuth values used in the test are 0°(frontal) / 30° (front-lateral) / 240° (back-lateral) / 180° (back). Of course, each of the samples that have to be repeated, and each of the sequences, must be reproduced at the same degrees of azimuth.

Given these rules, and applying a random criterion, the order of the samples was chosen as the following:

1. Distance slider 0, 0° azimuth
2. Distance slider 400, 30° azimuth
3. Distance slider 200, 0° azimuth
4. Distance slider 1000, 240° azimuth
5. Distance slider 700, 180° azimuth
6. Distance slider 200, 30° azimuth
7. Distance slider 250, 30° azimuth
8. Distance slider 300, 30° azimuth

9. Distance slider 350, 30° azimuth
10. Distance slider 0, 0° azimuth
11. Distance slider 50, 0° azimuth
12. Distance slider 100, 0° azimuth
13. Distance slider 150, 0° azimuth
14. Distance slider 700, 180° azimuth
15. Distance slider 400, 30° azimuth
16. Distance slider 1000, 240° azimuth
17. Distance slider 900, 240° azimuth
18. Distance slider 800, 240° azimuth
19. Distance slider 700, 240° azimuth
20. Distance slider 500, 30° azimuth.

8.1.5 Implementation of the testing platform

The testing platform was implemented using MaxMSP, creating a simple graphic interface (shown in Figure 2). A copy of the developed testing application can be found in the CD attached to this dissertation, *see* Appendix E. Within the platform there is an "instructions" page (a separate window appears once the appropriate button is pressed); an on-off button; a "comments section", where it is possible to insert a few words about the "relationship of the listener with music"; two VU meters; a "save the test" button, and the list of the listening samples. For each listening sample there is a button to start the playback of the sample, another to save the choices, and a checkbox in which to record the choices already made.

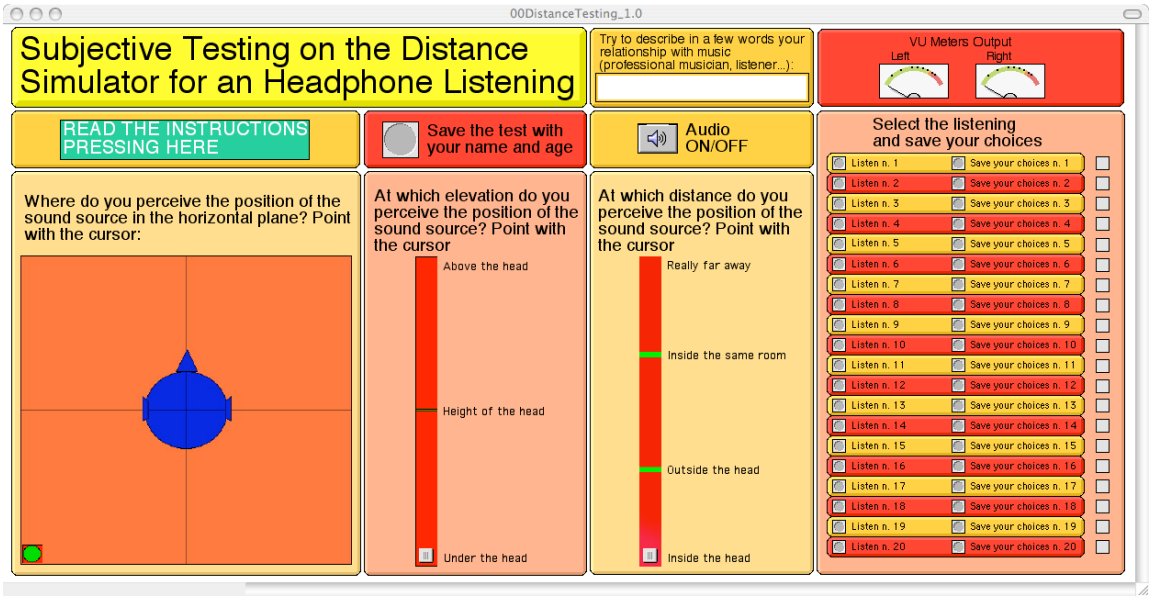

Figure 2. The Binaural Reverb and Distance Simulation testing platform.

8.1.6 Analysis of the results

All of the data from the test were collected, and different charts computed from the tables. Within the charts that follow, the different listened-to samples are displayed along the X axis, and the values of the answers of the corresponding listener along the Y axis, on a scale from 1 to 100. For certain charts, both the estimated and simulated values are displayed. Due to the fact that the goal of this test was to obtain some feedback and to certify the effectiveness of the theoretical background behind the distance simulation technique, no analysis of the results regarding azimuth and elevation localization data has been carried out.

In the following paragraphs the analysis of the data regarding the distance estimation, divided by listening sets, is reported.

Listening set 1-10 (Chart 1): same simulated distance, different answers. In the first and the tenth samples, the reproduced signal was the original mono soundfile (pure diotic stimulus). The estimation of the localization of the sound source, in this case, should have been exactly in the middle of the head of the subject.

In Chart 1 it may be remarked that the average of the answers in the first listening is around 25, and in the second around 16. The possible reason for this is that the concept of the IHL (*see* Chapter 5) is often not properly comprehensible. In other words, it is difficult for a subject to estimate that the sound source reproducing what s/he is hearing

is located in the middle of his/her head. After a first listening, the sound source may seem to be located in a frontal position. This is the same sensation that is present when music is listened through headphones: the sound seems to come from outside of the head. However, after the listener is presented with binaurally spatialized signals (in this case, after nine binaural stimuli), the perception of the distance is recalibrated, resulting in a closer (in terms of distance) estimation for a diotic stimulus. For this reason, the two judgements of the same sample, yet at two different stages of the test, are significantly different, and the value of the latter estimation of the distance is, as expected, lower than the former

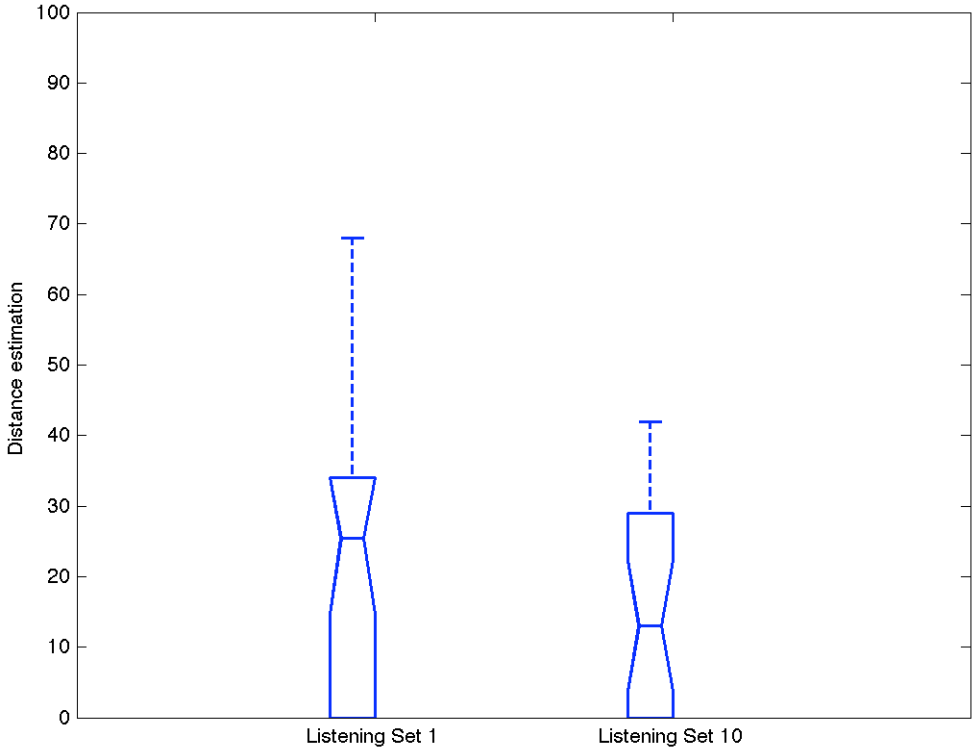

Chart 1. Boxplot (sample minimum, lower quartile, median, upper quartile and sample maximum, excluding the outliers) for the estimation of distance; Listening Set numbers 1 and 10.

Listening sets 10-13 (Chart 2): the distance is increased from 0 to 150 in increments of 50. Chart 2 shows how the average of the estimations of the distances increases from 16 to 34, following the increase of the "real" simulated distance. The passage between 0 and 150 is highly important because it corresponds with the cross-fade between the

original mono (non-spatialized) soundfile (distance slider value 0) and the spatialized (with anechoic, early reflections, and reverberant HRIRs and BRIRs) signal (distance slider value 200), which should mark the movement of the apparent image of the sound source from inside (IHL) to outside the head. These are particularly important data because they underline the fact that, even if the absolute evaluation of the distance is not particularly close to the "real" simulations, the increases in the distance simulation correspond to the increases in the distance perceived.

It is also important to notice that in Figure 1 the first two steps of the distance slider (on the left of the platform) are "inside the head" (value 0) and "outside of the head" (value 30): the answer to the tenth sample (distance slider value 0) is 16 (between the first two steps), and the answer to the thirteenth sample (distance 150) is 34 (above the second step), therefore this confirms that adding components of the signal spatialized using even only anechoic HRIRs can highly contribute to the apparent localization of virtual sound sources outside the head.

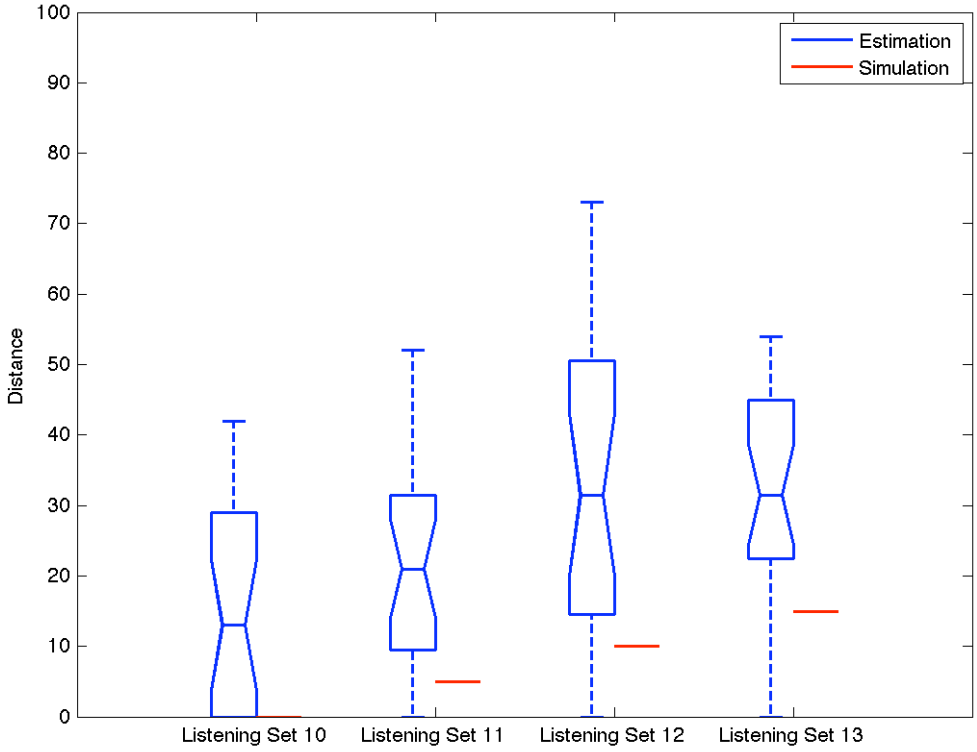

Chart 2. Boxplot (sample minimum, lower quartile, median, upper quartile and sample maximum, excluding the outliers) for the estimation of distance; Listening Set numbers 10, 11, 12, and 13.

Listening sets 16-19 (Chart 3): the distance is decreased from 1000 to 700 in increments of 100. Chart 3 shows how the average of the estimations of the distances decrease from 82 to 64, following the decrease of the "real" simulated distance. Even in this case the absolute estimation of the distance is irrelevant, while the fact that the subjects estimated a decrease of the distance, following the "real" simulation, is particularly relevant. Between 1000 and 700, the distance simulation is based on the increase of the direct-to-reverberant signal ratio and on the simulation of the frequency-dependent attenuation of the air. Within this listening set, it was highly important to realize that this kind of "simulation of long distances" works, even given the fact that the human hearing system has no precise ability to establish the distance of a sound source for long distances (*see* Chapter 5).

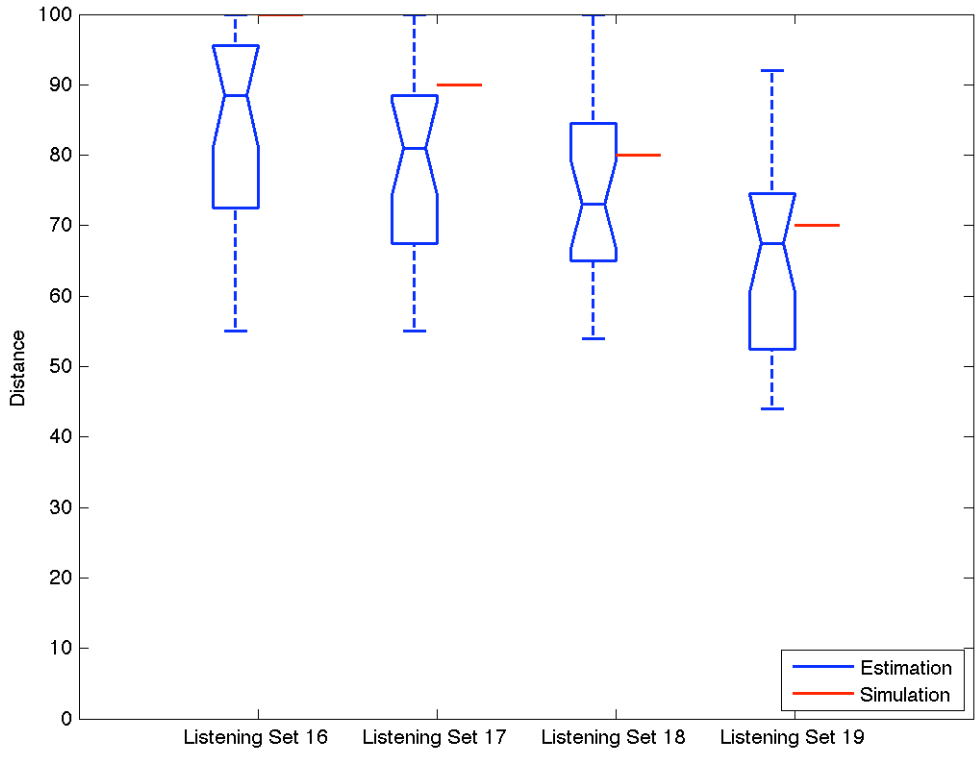

Chart 3. Boxplot (sample minimum, lower quartile, median, upper quartile and sample maximum, excluding the outliers) for the estimation of distance; Listening Set numbers 16, 17, 18, and 19.

Listening sets 6-9 (Chart 4): the distance is increased from 200 to 350 in increments of 50. Chart 4 demonstrates how the average of the estimation of the distances increases from 38 to 44, following the increase of the "real" simulated distance. In this case, the variation in the average of the answers between the sixth and ninth listening set is not large, although it is particularly relevant. In fact, from 200 to 400 there is the passage between the anechoically spatialized signal (with very few early reflections and reverberant components) and the early reflections one (at 400 anechoic and early reflections signals are mixed at –6 dBfs). The importance of these answers lies in the fact that the signal spatialized with early reflections HRIRs, and mixed with the anechoic HRIRs, seems to give a sensation of distance that helps in localizing sound sources outside of the head.

This is relevant considering also the binaural reverb technique, as the mix between direct, early reflections and reverb signals seems to work efficiently to the same extent as do the correspondent different sections in an algorithmic reverb design.

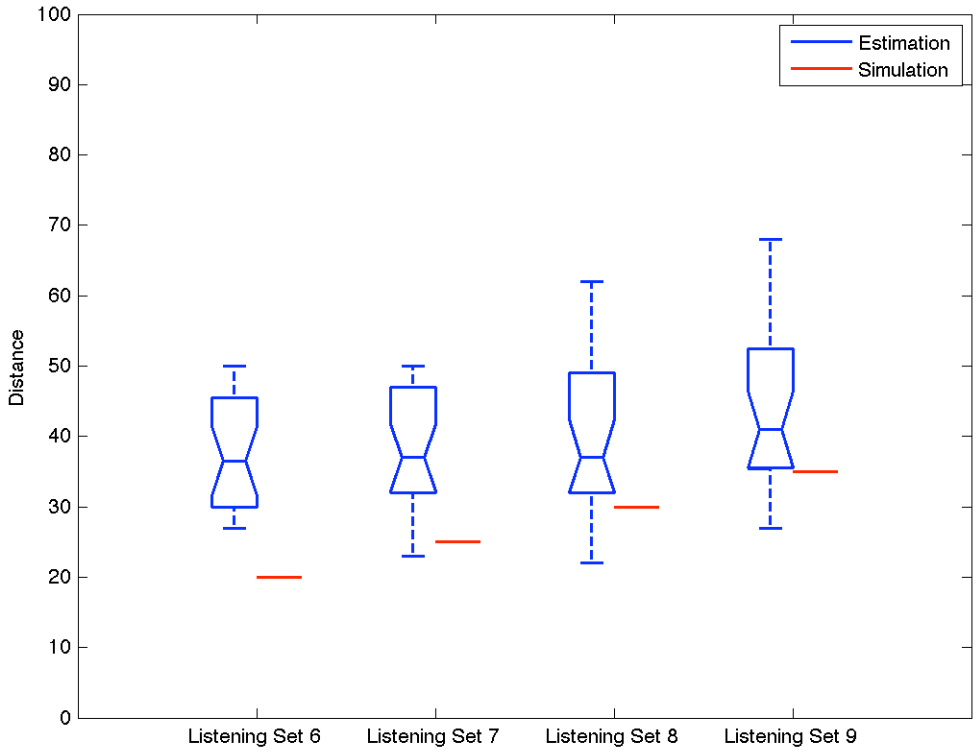

Chart 4. Boxplot (sample minimum, lower quartile, median, upper quartile and sample maximum, excluding the outliers) for the estimation of distance; Listening Set numbers 6, 7, 8, and 9.

8.1.7 Relevant issues

Two important issues have been outlined during the analysis of the data collected from this test:

- Variability of the results: it was decided to evaluate the average response of all of the subjects rather than the individual responses. For this reason, the charts show that the range of the sample answers is rather wide; nevertheless, it was considered sufficient for this early test to evaluate only the averages of the response, without the need to analyse and weight the single responses of every individual, or to analyse standard deviations, sample mean errors, etc. (the first lines of Section 8.1 offers a justification for this choice).
- The results are not close to the "right" answers (correspondent to the simulation values): it has not been important to test an absolute evaluation of the distance, simply because this is too subjective an estimation (see Chapter 5). Even in real, not "synthetic", situations, this estimation could be especially difficult. Furthermore, the values of the distance slider do not actually correspond to any measuring unit for quantifying the distance of a sound source. The important data are that the variations of the estimation values follow the variations of the "right" answers, thus the variations of the simulated values. When the distance is artificially (binaurally) increased, the average of the responses informs the tester that the estimated distance similarly increases, and the same correspondence occurs when the distance is reduced.

8.1.8 Conclusions

The evaluation of the results emerging from this first subjective perceptual test shows it to be positive. Even were the analysis not expanded in depth (which was itself not the aim of the test), it may be stated that the test confirms the effectiveness of the theoretical background on the basis of which the binaural reverb and distance simulation technique were developed. This was simply a first step, allowing the improvement of the algorithm, and its implementation inside a more complete binaural spatialization tool.

8.2 Binaural reverb simulation test

While the test described in Section 8.1 aimed to be a first evaluation of the distance simulation technique, this second perceptual test is performed on the final application (as described in Chapter 7). Its focus is on the main innovation of this PhD: the spatialization of signals using a weighted mix of the different HRIR and BRIR components, allowing environmental characterization through cross-synthesis processing between synthesized reverberant IRs and the measured BRIRs.

The perceptual test is based on a comparative task, an attempt to evaluate the quality of the technique developed as compared with other environmental simulation techniques, be they stereo, multichannel or binaural. The following sections outline the test objective, the planning, the implementation, and the analysis of the results.

8.2.1 Objective

The objective of this final perceptual test was to verify the perceived spatialization quality of the application developed as compared with other stereo, multichannel and binaural environmental simulation techniques, as well as to evaluate how realistic the effect generated by these simulations is. The same signals and the same environmental simulations were used in generating virtual environments using five different applications (including that developed within this research work); the quality and realism of these was compared, evaluated and, finally, ranked.

The expected results were that the use of measured BRIRs could significantly increase the realism and quality of the simulation, and that performing cross-synthesis operations on the BRIRs, despite offering an obvious amelioration of the flexibility within the simulation, did not influence the quality and realism of the spatialization.

8.2.2 Test planning

The testing platform has been implemented using MaxMSP; the developed platform, acting as a simple interface, is used for the playback of pre-processed signals and for the management and storage of the answers of the different subjects. A copy of the developed testing application can be found in the CD attached to this dissertation, *see* Appendix E. The actual processing of the signal is carried out in advance, offline, on the selected platforms according to each binaural engine used.

Five different binaural spatialization engines (and/or configurations) are tested (any reference to these engines in the following pages will be made reporting only the Eng1, Eng2, etc. naming convention):

- Eng1. The application developed within this research work, performing environmental simulation using BRIRs and characterization through cross-synthesis operations.
- Eng2. The direct path components are simulated using the anechoic HRIRs measured within this research work, while the environmental simulation is performed using a standard convolution stereo reverb (in this case, Audio Ease Altiverb 7).
- Eng3. The original idea was to use the IEM Pure Data binaural engine (*see* Noisternig, 2003a and 2003b, and Musil, 2005), based on the Ambisonic environmental simulation converted to binaural. Unfortunately, a stable Universal Binary version of this engine was not available when the test was being programmed; therefore another similar spatialization technique was chosen. Using the same CATT-Acoustic models programmed for Eng4, the simulation was done employing 2nd Order Ambisonics IR, converting them to binaural using an Ambisonic-to-binaural method based on virtual loudspeakers (twelve virtual loudspeakers placed at the angles of an icosahedron, *see* Section 7.2). For the binaural conversion, IRCAM Listen HRTF is used (subject number 1032, *see* Section 2.4.6).
- Eng4. Using the CATT-Acoustic acoustic simulation software, an acoustic model of a room was created and auralized directly in a binaural format.
- Eng5. A 5.1 acoustic room reconstruction (performed using Logic Audio Pro 8 from Apple and Audio Ease Altiverb 7), converted to binaural using the MIT HRIR database (*see* Gardner, 1994).

For each of the binaural engines cited above, three acoustic simulations are performed:

- Env1: A small dry room (dimensions 3m x 3m x 2.3m, RT60 ~0.4 sec +- 0.05 sec on the entire frequency range).
- Env2: A small-medium reverberant room (dimensions 4m x 6m x 2.5m, RT60 ~1 sec for the low frequencies, ~1.2 sec for the middle frequencies and ~0.9 sec for the high frequencies).

- Env3: A large reverberant room (dimensions 12m x 8m x 3.5m, RT60 ~1.6 sec for the low frequencies, ~1.9 sec for the middle frequencies and ~1.35 sec for the high frequencies).

For each acoustic simulation and each engine, thirty-six positions are simulated at different azimuth degrees (one for each 10 degrees of azimuth, from 0° to 350°) at a distance of one metre. The elevation parameter will not be tested within this experiment (the reasons for this are the same as those for the previous test, *see* Section 8.1).

The signals used are:

- Samp1: Female Speech (English, from the "Music for Archimedes" anechoic recordings, *see* Hansen, 1991)
- Samp2: Guitar music (Bach, from the "Music for Archimedes" anechoic recordings, *see* Hansen, 1991)
- Samp3: Synthesized clicks (eight repetitions, base frequency 440 Hz, duration 200 ms).

The test is carried out in this way:

- The subject hears, sequentially, the same sample (same signal, same azimuth position, same simulated environment) spatialized using two different engines, and performs his/her choices regarding the listened-to sample.
- Sixty times, the listener is presented with the pair of stimuli: each time, the simulation engines, the environment, the azimuth position and the signal are changed.
- The following rules are followed for the extraction of the list of the sixty pairs:
    - The numbers of the total of possible pairs is given by the combinations of the five engines in groups of two. There are therefore ten combinations (binomial coefficient), and every possible pair is presented six times in a random order.
    - With three different environmental simulations, for the six repetitions of the same pair of the previous point each environmental simulation is presented twice, in a random order.
    - With three different audio samples, for the six repetitions of the same pair each sample is presented twice, in a random order.
    - The azimuth degree is randomly chosen for each of the pairs; it was chosen to be the same for both the elements of the pair. The randomization process for the azimuth selection has a "non-repetition" rule: for the first thirty-six stimuli, all

thirty-six azimuth positions are presented in random order. In the remaining twenty-four, twenty-four azimuth positions, with no repetition, are presented in a random order.

- Therefore, each possible pair is presented twice for each of the three simulated environments, for each of the three signals, pseudo-randomizing the azimuth selection (the same for each element of each pair).

Two questions are asked of the subjects (the five options for replies are shown in brackets):

- *Which of the two examples sounds more real? (*surely the first – could be the first – I don't know – could be the second – surely the second*)*
- *Which of the two examples do you like more? (*surely the first – could be the first – I don't know – could be the second – surely the second*)*.

In order to obtain significant results, the test has to be carried out with a minimum of fifteen to twenty subjects.

8.2.3 Tested room simulations

As already stated in Section 8.2.2, all signals were pre-processed using different applications. The parameters and configuration used for the processing are reported here.

Eng1: developed application

The simulation of the three environments is carried out as reported in Chapter 7 (*see* Section 7.3.4). Regarding the large environment, the BRIRs are left unprocessed, due to the fact that the requirements in terms of dimensions and RT60 for the environment to be simulated (*see* Section 8.2.2) are very similar to the actual characteristics of the room where the BRIRs were measured. This factor will be of particular importance when the analysis of the results is made.

Eng2: HRIRs with stereophonic reverb

The binaural simulation is carried out using the anechoic HRIRs (see Chapter 4), while the environment is simulated employing a stereophonic version of the Audio Ease Altiverb 6 IR convolution reverb, using the same configurations and calibrations as those described for Eng5.

Eng3: CATT-Acoustic model with 2nd Order Ambisonic-to-binaural conversion

The simulation of the three environments is performed using three acoustic models realized with the CATT-Acoustic simulation software, and 2nd Order Ambisonics impulse responses are calculated for each of the possible source positions (from 0° to 350° of azimuth, 0° Elevation). The IRs are then converted to binaural using an icosahedron virtual loudspeakers array and the anechoic HRIRs measured within this research (for more information about this approach, see Chapter 7), and convolved with the three signals to be spatialized. It is important to underline that the geometrical characteristics, as well as the RT60 for the different frequency bands, are the same as those of the models created for the BRIR characterization process (see Section 7.3.4)

In Figure 3 the 2D and 3D views of the large room (Env3) can be seen as an example. The plans show the position of the head and the circles corresponding to the calculated source positions. Charts 5, 6 and 7 show the RT60 diagrams for the three simulated environments.

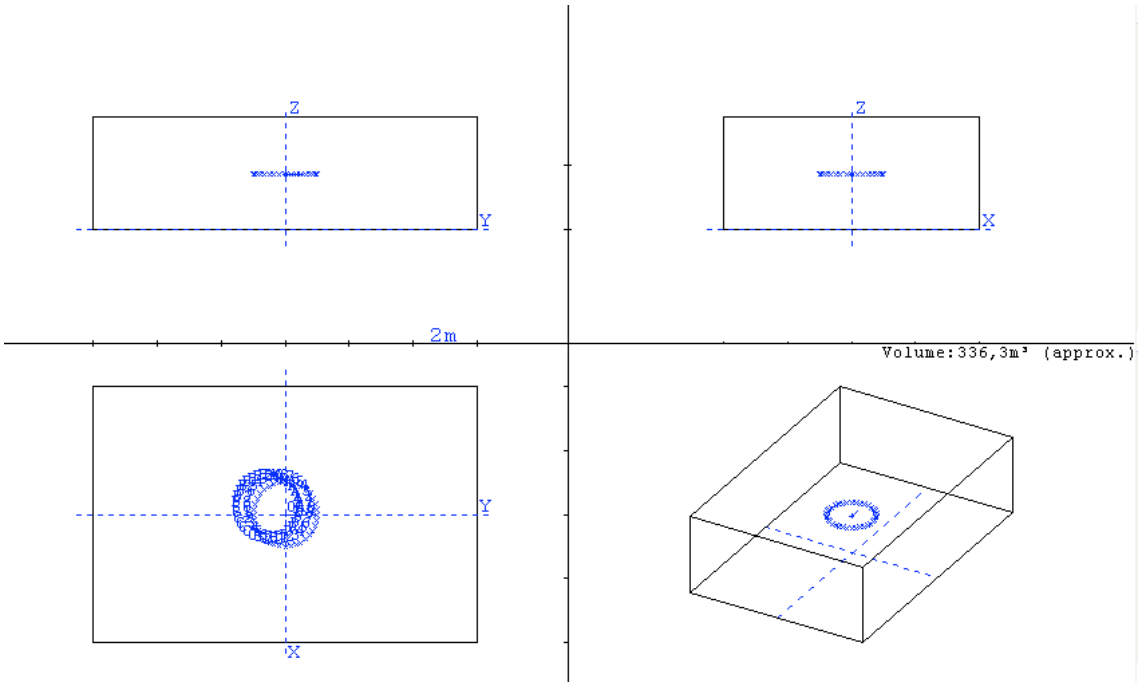

Figure 3. 2D and 3D view of the large room (Env3) model created with CATT Acoustics.

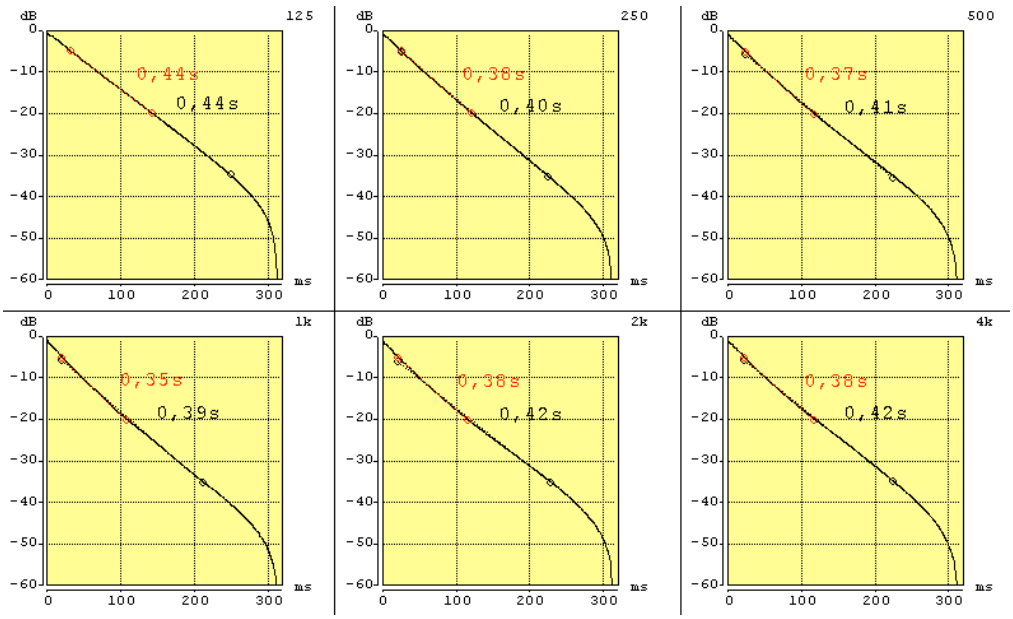

Chart 5. RT60 for six frequency bands of the small room model (Env1).

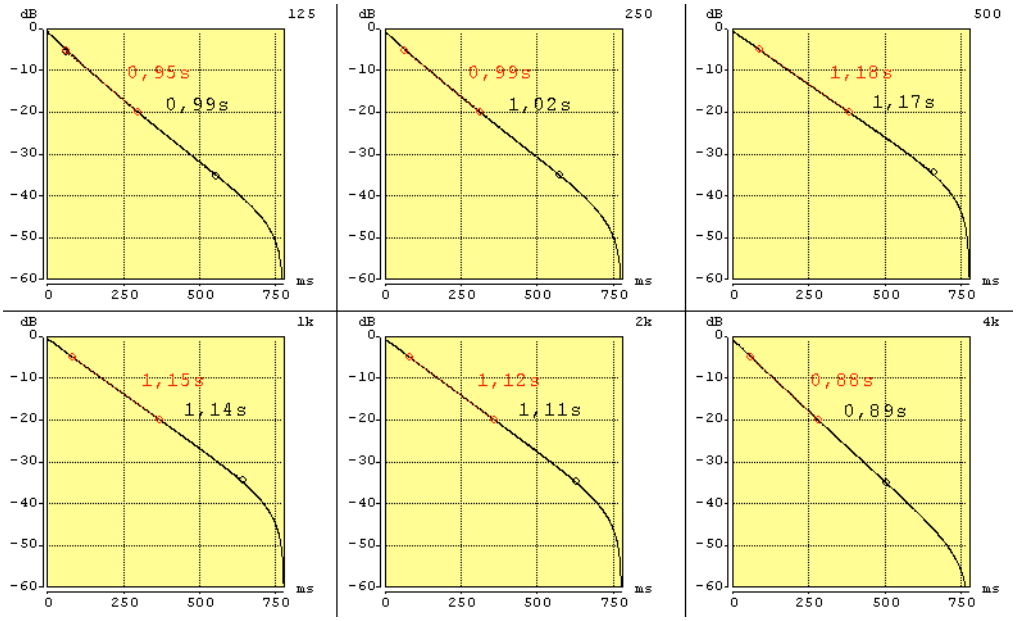

Chart 6. RT60 for six frequency bands of the medium room model (Env2).

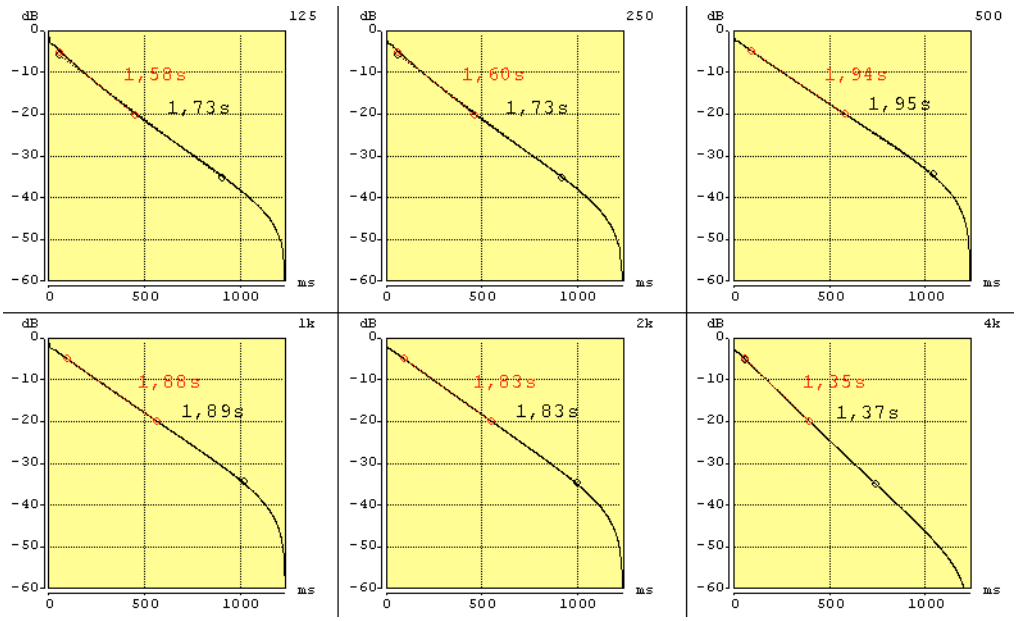

Chart 7. RT60 for six frequency bands of the large room model (Env3).

Eng4: CATT-Acoustic model with binaural auralization

The simulation of the three environments is performed using three acoustic models realized with the CATT-Acoustic simulation software, using exactly the same parameters and configuration as for Eng3, although the IRs are calculated directly in the binaural format using the binaural algorithm of CATT-Acoustics.

Eng5: 5.1 reverb, converted into binaural

The software Logic Audio Pro 8 with the Audio Ease Altiverb 6 IR digital reverb plugin is used for this simulation. For each position, the "diversity" parameter within the panning window, which alters the amount of audio distributed to the surround speakers, has been set to 0, so that the source is simulated as a point source. Thus, the direct signal comes only from the two nearest loudspeakers on the left and right sides of the source, while the reverb obviously comes from all five channels.

The 5.1 Surround Standard used is the ITU 775 (following this standard, the loudspeakers are positioned at: C 0°, L 30°, R -30°, LS 120°, RS -120°). The parameters and configurations used for the three environmental simulations are listed here.

Env1, small room:

- The IR was chosen from PostProductionAmbiences/DomesticBedrooms/Bedroom07, recorded in QUAD then converted in 5.1 by Audio Ease
- RT60 0.45 sec
- LowDamp 100% / Cross 300 Hz / MidDamp 100% / Cross 1000 Hz / HiDamp 100%
- Direct 0 dB / Colour 1.00 / EarlyRefl 0 dB / Delay 0 ms / Tail 0 dB / Delay 0 ms
- Input 0 dB / Output 0 dB / Front 0 dB / Rear 0 dB / Centre Bleed 0 dB / Mix 50%
- Equalization: ALL FLAT (0 dB over the entire frequency range).

Env2, medium room:

- The IR was chosen from PostProductionAmbiences/DomesticBedrooms/Bedroom04, recorded in QUAD then converted in 5.1 by Audio Ease
- RT60 1.22 sec
- LowDamp 80% / Cross 300 Hz / MidDamp 95% / Cross 1000 Hz / HiDamp 75%
- Direct 0 dB / Colour 1.00 / EarlyRefl 0 dB / Delay 0 ms / Tail 0 dB / Delay 0 ms
- Input 0 dB / Output 0 dB / Front 0 dB / Rear 0 dB / Centre Bleed 0 dB / Mix 50%
- Equalization: ALL FLAT (0 dB over the entire frequency range).

Env3, large room:

- The IR was chosen from RecordingStudios/AllaireNeveRoom, recorded in QUAD then converted in 5.1 by Audio Ease
- RT60 1.61 sec
- LowDamp 100% / Cross 300 Hz / MidDamp 115% / Cross 1000 Hz / HiDamp 85%
- Direct 0 dB / Colour 1.00 / EarlyRefl 0 dB / Delay 0 ms / Tail 0 dB / Delay 0 ms
- Input 0 dB / Output 0 dB / Front 0 dB / Rear 0 dB / Centre Bleed 0 dB / Mix 50%
- Equalization: ALL FLAT (0 dB over the entire frequency range).

8.2.4 Analysis of the results

Twenty subjects performed the test, and no major problem was found during the performance.

Within the analysis of the data, for each pair comparison a score is associated with the individual engine in the following way, with a score given for each of the two questions posed to the different subjects:

if the engine is evaluated as being "*surely*" the better of the two, 2 points are given, and -2 to the other option. If the choice is "*could be*", then only 1 point for the first and -1 for the second. If the choice is a "*I don't know*", 0 points are given to both signals. The score of the whole test for each engine is then averaged for each individual, and from this result emerge the data reported in the following charts.

Chart 8 shows the global score for all engines relative to the two questions. The most significant data extracted from the diagram are that Eng1 (the application developed within this research work) was consistently considered most successful while Eng2 (the anechoic binaural spatialization with stereo reverb) was considered the least successful. The two CATT-Acoustics simulations (Eng3 performed in the 2nd Order Ambisonics domain converted to binaural, and Eng4 directly in the binaural domain), understandably, were evaluated as being very similar, and Eng5 (5.1 reverb converted to binaural) is ranked in the second position yet far below Eng1.

A few considerations may now be offered:

- It can be considered a success that the application developed within this research work was evaluated as being the best over five proposed applications. Another very important item of data gathered from observing the minimum and the maximum values for each box is that for no subject did Eng1 achieve an average score of less than 0 (excluding the outliers) for both questions. Compare this with the result that Eng2 never reached a score greater than 0.
- The differences between the scores for the answers to the two questions cannot be considered relevant to any of the engines in such a global evaluation. Nevertheless, through commenting on the other diagrams reported in this section a closer analysis will be made of the topic.
- The very similar scores for Eng3 and Eng4 are explainable given the fact that the two simulations come from the same virtual acoustics models, generated using the same software. It is not known how CATT-Acoustics actually performs the binaural rendering of the calculated IR (Eng4), yet, considering these results, it may be stated

that this method is probably not too dissimilar to that used for the conversion between 2nd Order Ambisonics and binaural performed for Eng3.

- The weak scores for Eng2 are explainable given the fact that the environmental simulation was performed stereophonically (in one dimension), while the direct path component was processed with anechoic HRIRs; the mix of binaurally spatialized signals with non-spatialized seems to generate a negative effect in terms of both the quality and the realism of the simulation. This is a further confirmation of the fact that while performing 3D binaural spatialization, the environmental simulation, too, needs to be performed in 3D, possibly directly in the binaural domain. Many algorithms working similarly to Eng2 have been found and negatively evaluated in Chapter 2.
- It may be stated that, except for the stereo convolution reverb of Eng2, environmental simulation based on convolution (Eng1 and Eng5) was preferred to the algorithmically generated one (Eng3 and Eng4), confirming the fact that convolution reverbs tend to sound more real than algorithmic ones (*see* also the introduction to this research work in Chapter 0).

In Charts 9-12, the scores are reported in boxplot (*see* the captions) diagrams, separately for each of the three simulated environments and for each of the three signals; again, individually for each of the two questions. It should be noted that in these diagrams, too, the ranking of the five engines is the same as in Chart 8. It proves that, to a certain extent, the differences in the evaluations are due only to the different engines, and that the different signals and different simulated environments do not influence the listeners' judgment.

An interesting result comes from the analysis of the data within Charts 9 and 10, which correspond to the answers to the two questions separately for each of the three simulated environments). The ranking is calculated considering the individual averages of the three simulated environments and is the same as the global reach of Chart 8; however, within different engines differences between the different environments may be outlined. A particularly interesting element emerges from the Eng1 data. In the simulation of the large environment, the BRIRs were left unprocessed (*see* Section 8.2.3), while for the simulation of the medium and small environments cross-synthesis processing was carried out. The values in Chart 9 show that the average score for the large

environment is noticeably smaller than that for the other two environments, even if it is never lower than any of the average scores for Eng5. Thus, it may be stated that the cross-synthesis process did not affect the localization cues within the BRIRs, as had been feared (*see* Section 6.4.2); the localization quality and performances were left unaffected. However, the increase in the average score for the small and medium environments cannot be explained through this process. This constitutes possibly one of the most substantial results of the test: the BRIR characterization process seems to be proved to work, allowing the realism and spatialization quality of a convolution-based environmental simulation, and with the flexibility of an algorithmic method.

Particularly relevant is the fact that the ranking of each environment simulation within the individual engines is not always the same, strengthening the assumption that the score depends not on the typology of the environment, but only on the different engines.

Similar observations can be drawn from Charts 11 and 12, relative to the individual scores of the three different signals. Here, though, for the speech signal the top score is given to Eng5 rather than Eng1, for both of the questions. It is nevertheless true that for a global ranking the relevant data are those in Chart 8, corresponding to the average of all of the different individual environments and signal data.

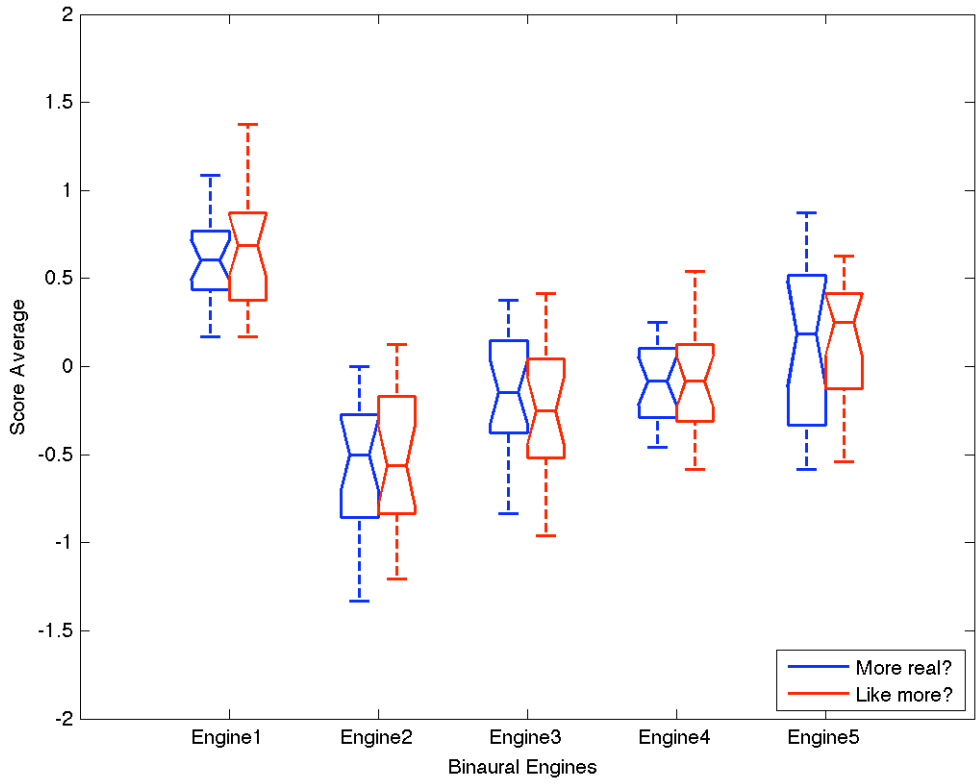

Chart 8. Boxplot (sample minimum, lower quartile, median, upper quartile and sample maximum, excluding the outliers) relative to the answers for all of the environments and signals.

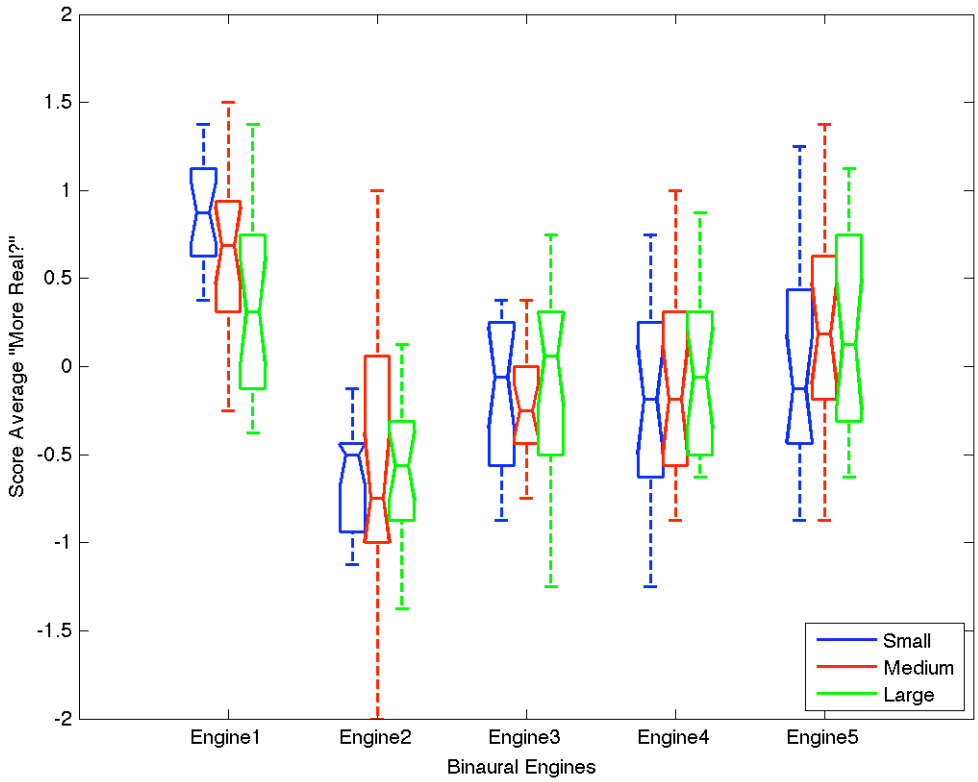

Chart 9. Boxplot (sample minimum, lower quartile, median, upper quartile and sample maximum, excluding the outliers) relative to the answer number 1 (more real) for the three environments (all of the signals).

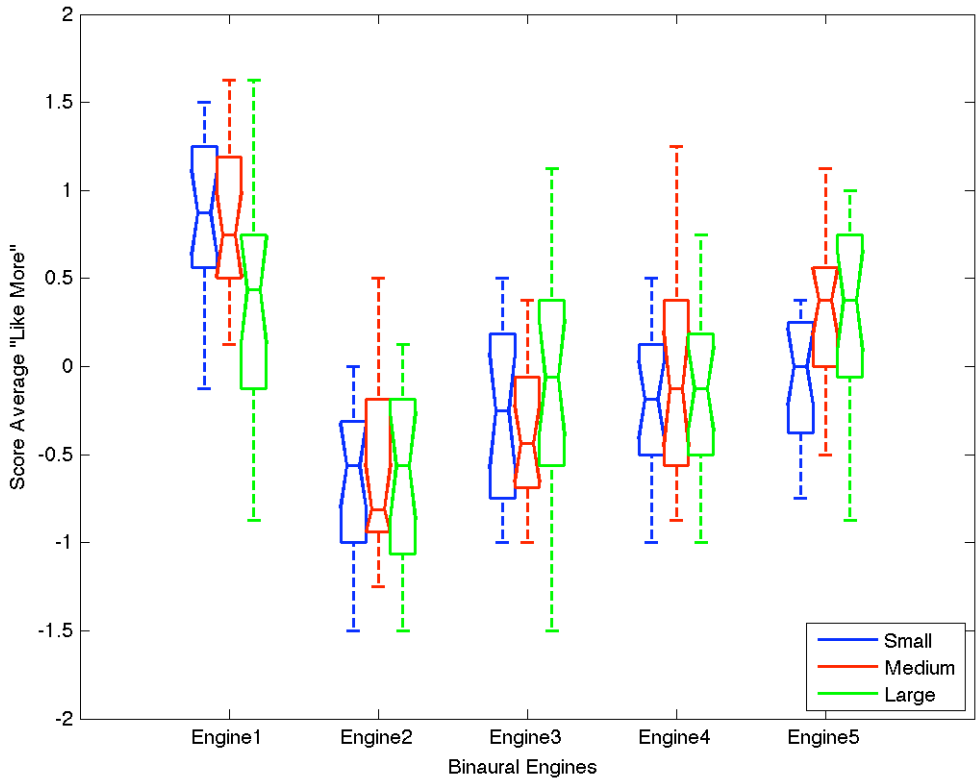

Chart 10. Boxplot (sample minimum, lower quartile, median, upper quartile and sample maximum, excluding the outliers) relative to the answer number 2 (like more) for the three environments (all of the signals).

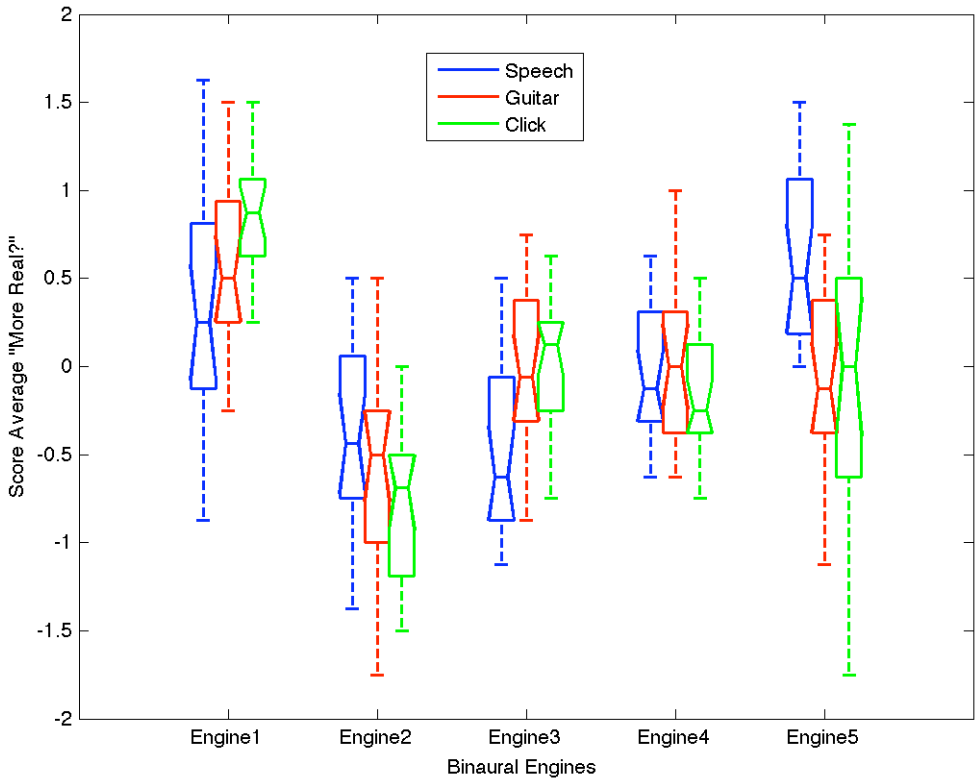

Chart 11. Boxplot (sample minimum, lower quartile, median, upper quartile and sample maximum, excluding the outliers) relative to the answer number 1 (more real) for the three signals (all of the environments).

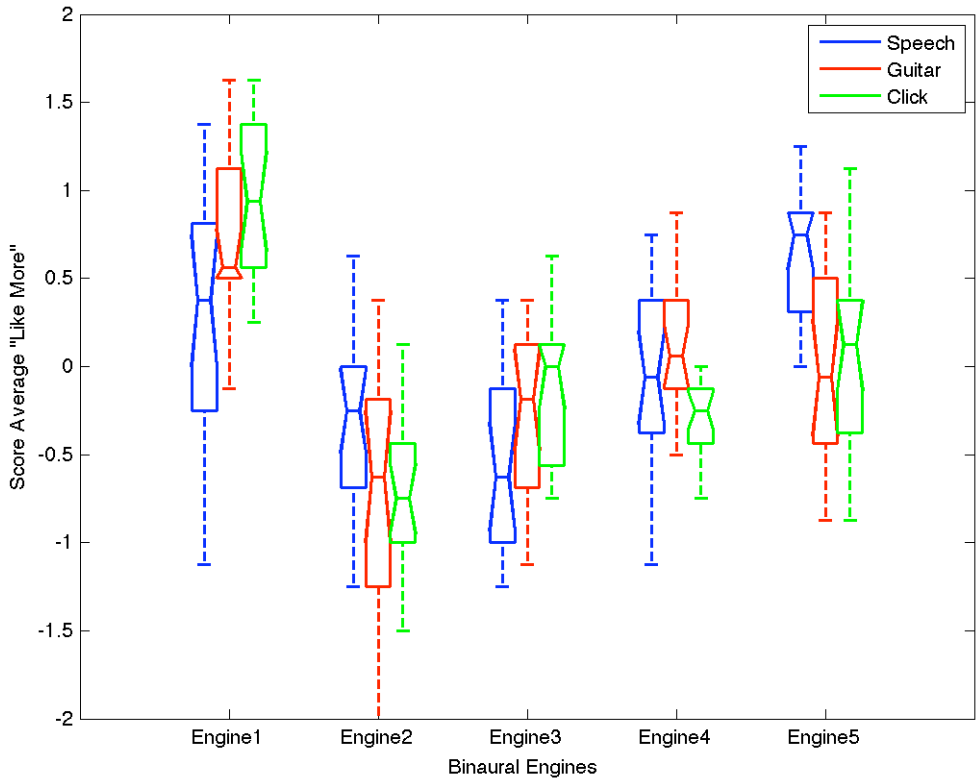

Chart 12. Boxplot (sample minimum, lower quartile, median, upper quartile and sample maximum, excluding the outliers) relative to the answer number 2 (like more) for the three signals (all of the environments).

In Charts 13 and 14, the mean of the scores have been plotted with error bars reporting both the standard deviation and the unbiased estimation of the standard deviation of the sample mean, calculated by dividing the sample standard deviation by the square root of the sample size (in this case, the square root of twenty). These results offer an estimation of the possible results if this test were carried out again with different subjects. The data reported in the two charts show how an overlap of results in terms of standard deviation of the sample mean, therefore a possible different ranking of the five engines performing the test again on a different subject population, is possible only between the third and the fourth engines, but not between any of the others. Above, it was already noted how the similarities in terms of processing between these two engines showed a very similar result in terms of average score; therefore, this overlap should not be considered as a relevant parameter.

Furthermore, Table 1 shows the distance in terms of standard deviations (including the worst case in terms of standard deviation of the sample mean) between the first engine and the others for both of the answers. Except for the comparison with the fifth engine, these distances are nearly always greater than two standard deviations; given the worst case and therefore subtracting from the mean the standard deviation of the population sample, the distance remains relevant. This means that the ranking may be considered robust and replicable, and that the results may be considered statistically significant.

In the case of the comparison with the fifth engine, only a distance of approximately one standard deviation could be found. This does not mean that the results are not at all robust and replicable, but simply that in terms of significance, the comparison between the first and the last engines gives no result as relevant as in the comparison between the first and the other engines.

It is interesting to note that between the data relative to the two answers some differences are found in terms of standard deviations distance. This can be explained through the fact that the standard deviation for the data supplied by the first answer is smaller than that for the data supplied by the second answer. Nevertheless, in both cases the assumptions made in the previous paragraph may be considered to hold.

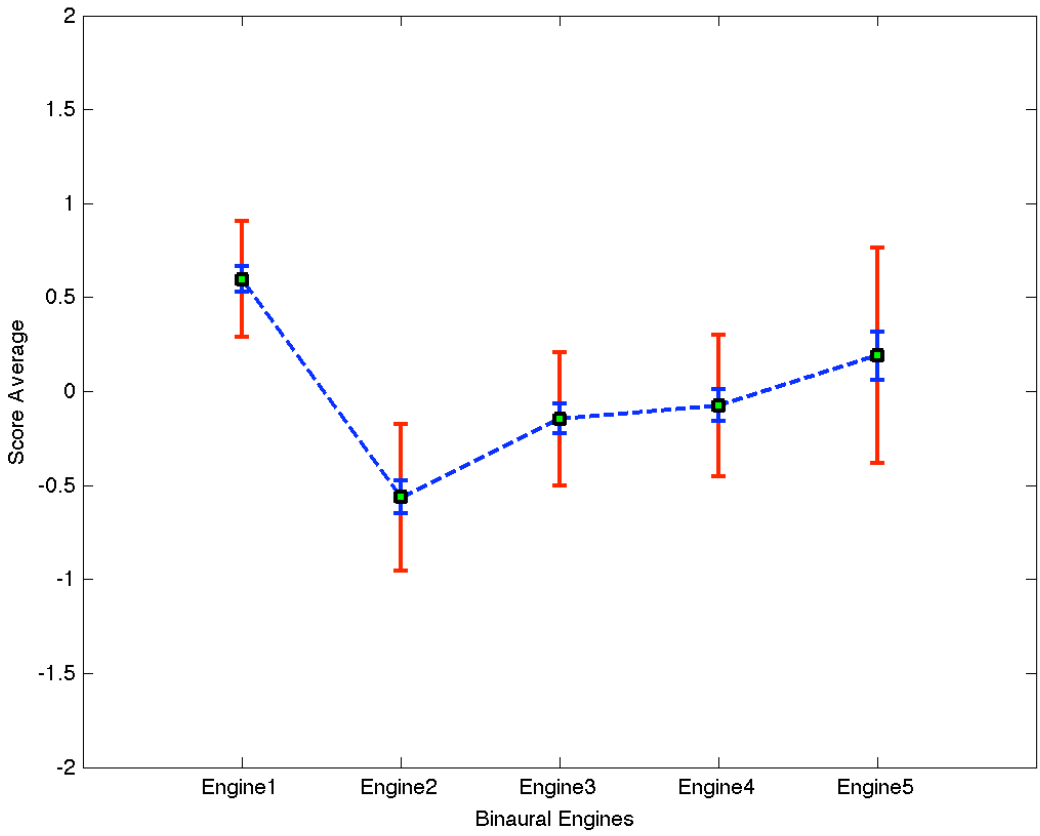

Chart 13. Average scores for each of the engines relative to the answer number 1 (more real). The error bars report the unbiased estimate of the standard deviation of the sample mean (sample standard deviation divided by the square root of the sample size) in blue, and the standard deviation in red.

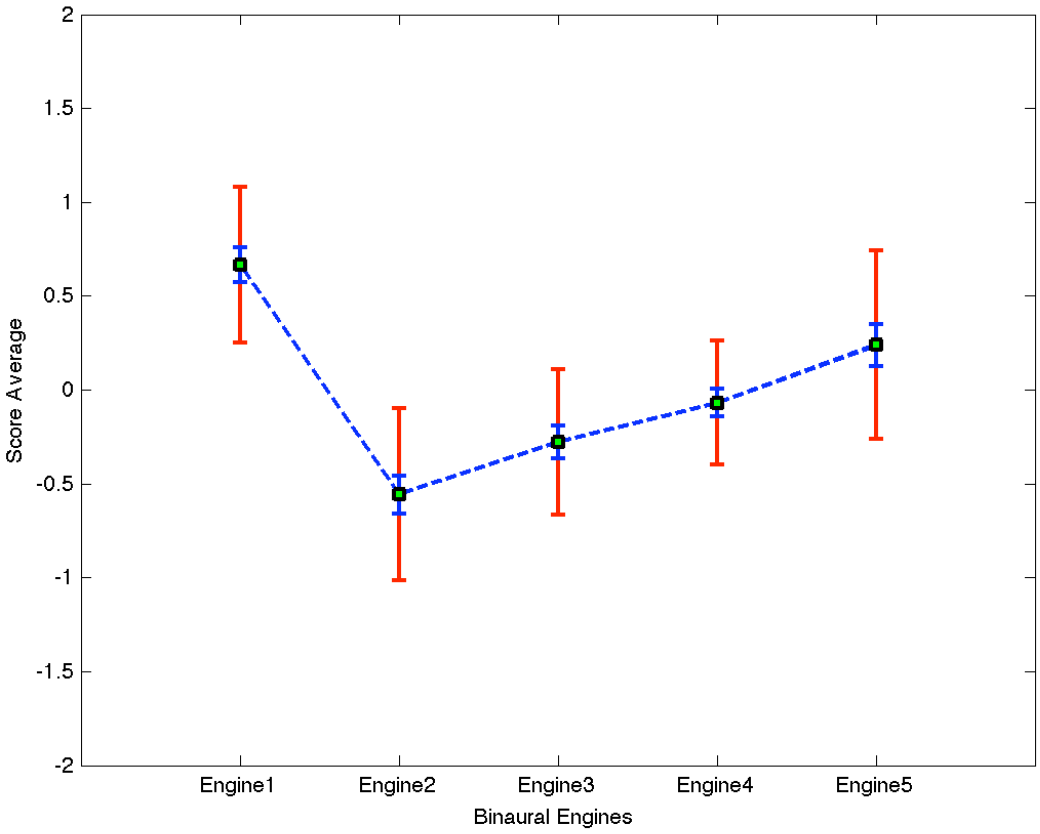

Chart 14. Average scores for each of the engines relative to the answer number 2 (like more). The error bars report the unbiased estimate of the standard deviation of the sample mean (sample standard deviation divided by the square root of the sample size) in blue, and the standard deviation in red.

Answer 1 (more real)				
	Eng1-2	Eng1-3	Eng1-4	Eng1-5
Number of StDev	3.760	2.403	2.174	1.316
Number of StDev (considering sample StDev)	3.537	2.180	1.950	1.093
Answer 2 (like more)				
Number of StDev	2.950	2.271	1.769	1.025
Number of StDev (considering sample StDev)	2.726	2.048	1.545	0.802

Table 1. The distance in terms of standard deviations (considering also the standard deviation of the sample mean) between the first engine and the others, for both answers.

In Charts 15 and 16 the score means of each of the engines are plotted in individual lines for each subject (relative to the two answers). At a first sight, the global image, the average, of the different lines perfectly fits with their actual ranking in terms of mean scores for the five engines. However, while it may be remarked how consistently the first and second engines have been scored (with nearly parallel lines), between the second and the fifth engines considerable amounts of crossing are found on the different lines. The first and the second engines have obviously been scored consistently because they represent the best and the worst cases within the sample, yet it is more difficult to explain the crossings between the score lines for the other environments. A possible explanation could be that the third, fourth and fifth engines (mostly, the third and the fourth) have been awarded very similar mean scores, thus it is understandable that different subjects ranked them in different positions above or below the mean, although always between the worst- and the best-evaluated engines, i.e., the first and the second. Nevertheless, the global shape of the lines confirms the results outlined in the previous paragraphs.

These two diagrams outline a difference in the score ranges of the different subjects between the first and the second answers. The answer to the first question (more real) generally gave a more compact score than the answer to the second question (like more). This could be explained through the formulation of the questions themselves. Asking which of the signal sounds more real could be considered a rather more objective question than which of the signals is better liked, and this could be the reason why the variance between the different subjects is larger for the second answer despite the fact that the mean remains approximately the same.

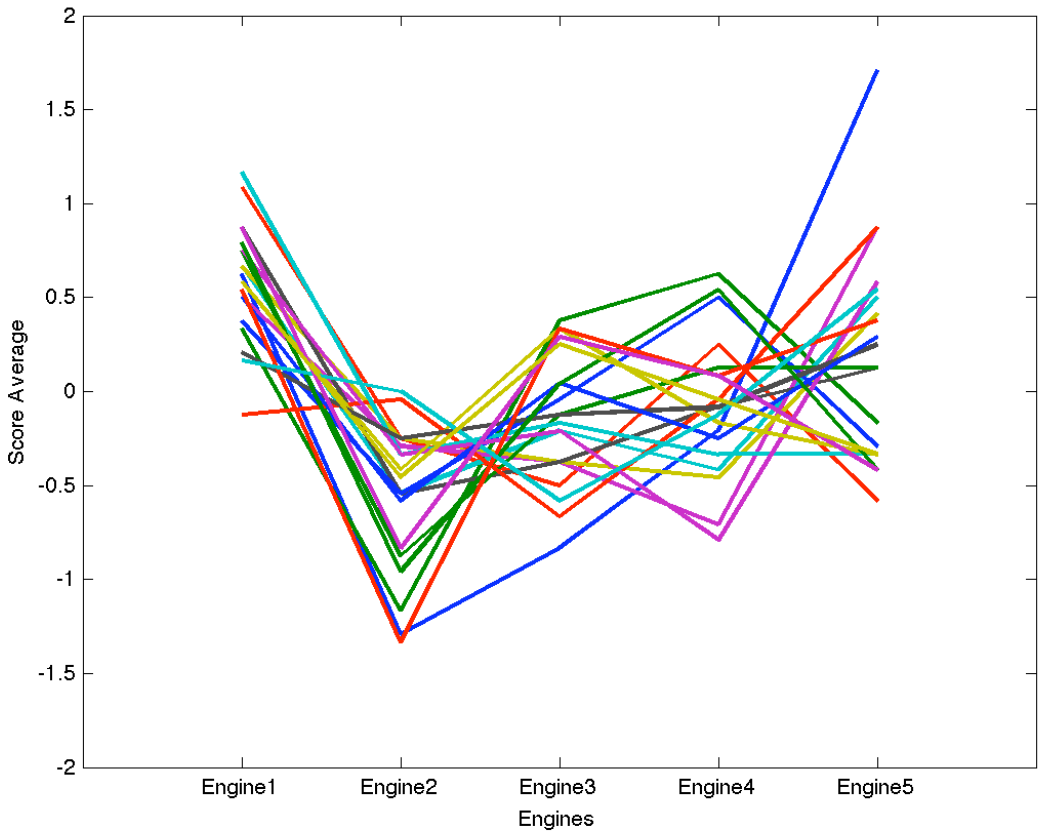

Chart 15. Score means of each of the engines plotted for each subject, relative to answer number 1 (more real).

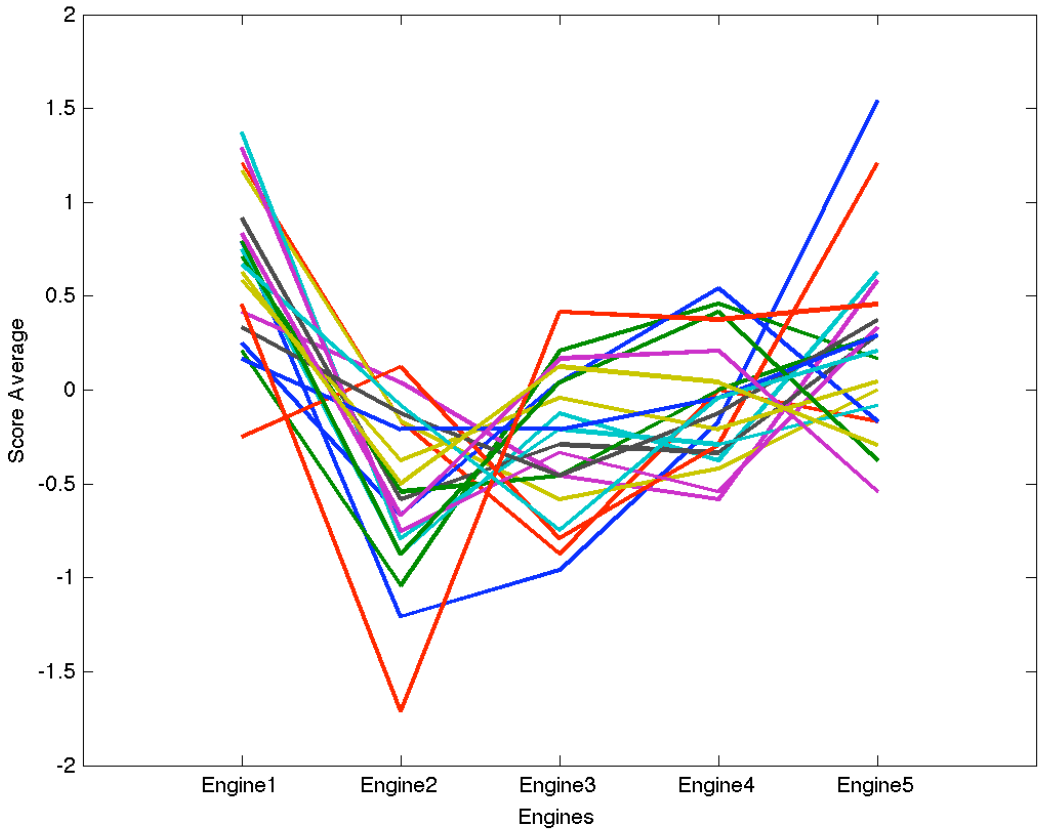

Chart 16. Score means of each of the engines plotted for each subject, relative to answer number 2 (like more).

In Charts 17 and 18, the curves of the score means for the five engines and for each of the subjects are plotted relative to the answers to the two questions. Analysing these diagrams shows how the blue line (corresponding to the first engine, that developed within this research work) is often (on fifteen occasions out of twenty) above the others; the green line (corresponding to the second engine, the binaural signal with stereophonic reverb) is often below. Therefore, for only five subjects out of twenty for both questions has the first engine *not* been ranked in the first position; out of these five individuals, for only one has it been ranked in the third position, and for the other four in the second position.

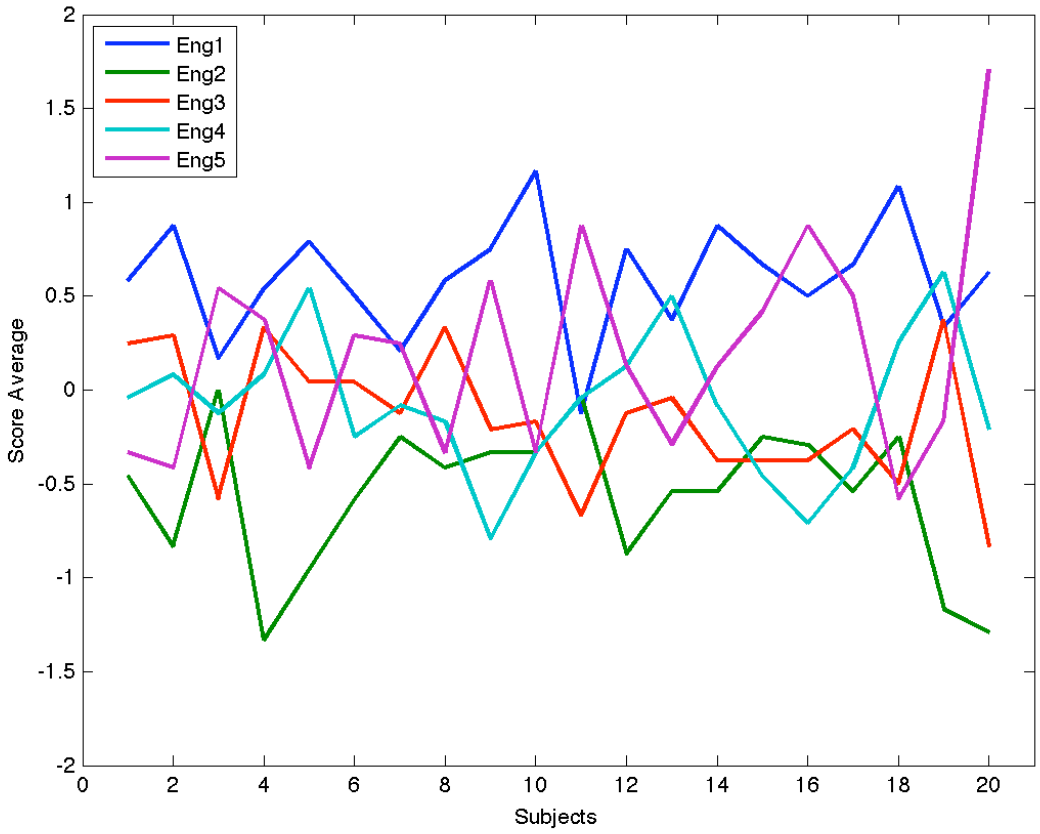

Chart 17. Score means for each of the subjects relative to the answer number 1 (more real).

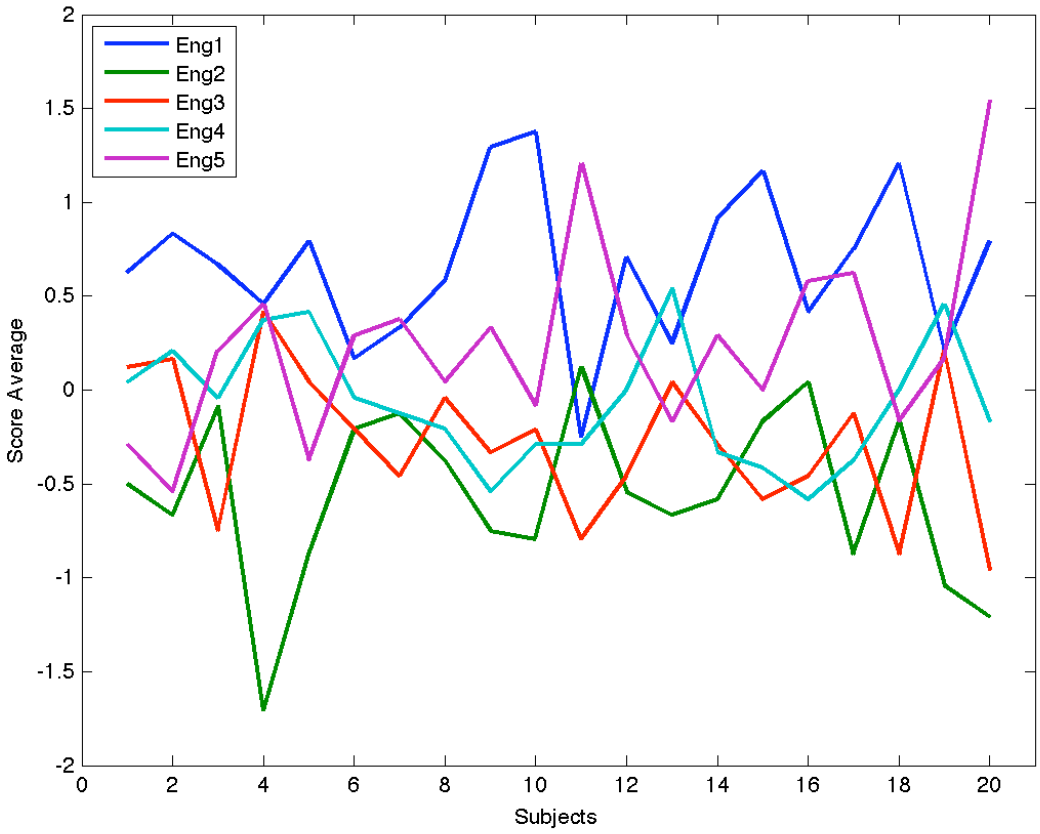

Chart 18. Score means for each of the subjects relative to the answer number 2 (like more).

Finally, all of the individual values of the means, standard deviation and inter-quartile ranges were collated in three tables (Tables 2, 3 and 4).

Total Data						
		Eng1	Eng2	Eng3	Eng4	Eng5
MORE REAL	Mean	0.596	-0.565	-0.146	-0.075	0.190
	StDev	0.309	0.391	0.354	0.377	0.573
	IQR	0.333	0.583	0.521	0.396	0.854
LIKE MORE	Mean	0.665	-0.558	-0.277	-0.069	0.240
	StDev	0.415	0.459	0.389	0.331	0.502
	IQR	0.500	0.667	0.563	0.438	0.542

Table 2. The mean, standard deviation and inter-quartile values for all of the environments and signals.

Mean						
		Eng1	Eng2	Eng3	Eng4	Eng5
MORE REAL	Small	0.850	-0.569	-0.150	-0.113	-0.019
	Medium	0.594	-0.525	-0.244	-0.063	0.238
	Large	0.344	-0.600	-0.044	-0.050	0.350
LIKE MORE	Small	0.838	-0.506	-0.256	-0.025	-0.050
	Medium	0.788	-0.606	-0.425	-0.069	0.313
	Large	0.369	-0.563	-0.150	-0.113	0.456
Standard Deviation						
		Eng1	Eng2	Eng3	Eng4	Eng5
MORE REAL	Small	0.444	0.595	0.430	0.735	0.598
	Medium	0.511	0.733	0.428	0.612	0.582
	Large	0.513	0.569	0.672	0.456	0.887
LIKE MORE	Small	0.426	0.713	0.497	0.647	0.461
	Medium	0.471	0.706	0.454	0.622	0.502
	Large	0.701	0.606	0.684	0.524	0.840
Inter-quartile Range						
		Eng1	Eng2	Eng3	Eng4	Eng5
MORE REAL	Small	0.500	0.500	0.813	0.875	0.875
	Medium	0.625	1.063	0.438	0.875	0.813
	Large	0.875	0.563	0.813	0.813	1.063
LIKE MORE	Small	0.688	0.688	0.938	0.625	0.625
	Medium	0.688	0.750	0.625	0.938	0.563
	Large	0.875	0.875	0.938	0.688	0.813

Table 3. The mean, standard deviation and inter-quartile values for the three simulated environments.

Mean						
		Eng1	Eng2	Eng3	Eng4	Eng5
MORE REAL	Speech	0.319	-0.344	-0.463	-0.094	0.581
	Guitar	0.588	-0.575	-0.063	0.025	0.025
	Click	0.881	-0.775	0.088	-0.156	-0.038
LIKE MORE	Speech	0.288	-0.275	-0.531	-0.056	0.575
	Guitar	0.744	-0.713	-0.213	0.119	0.063
	Click	0.963	-0.688	-0.088	-0.269	0.081

Standard Deviation						
		Eng1	Eng2	Eng3	Eng4	Eng5
MORE REAL	Speech	0.625	0.550	0.512	0.551	0.603
	Guitar	0.424	0.540	0.488	0.469	0.792
	Click	0.526	0.421	0.468	0.360	0.838
LIKE MORE	Speech	0.697	0.572	0.516	0.501	0.622
	Guitar	0.515	0.643	0.544	0.452	0.662
	Click	0.639	0.546	0.494	0.385	0.754

Inter-quartile Range						
		Eng1	Eng2	Eng3	Eng4	Eng5
MORE REAL	Speech	0.938	0.813	0.813	0.625	0.875
	Guitar	0.688	0.750	0.688	0.688	0.750
	Click	0.438	0.688	0.500	0.500	1.125
LIKE MORE	Speech	1.063	0.688	0.875	0.750	0.563
	Guitar	0.625	1.063	0.813	0.500	0.938
	Click	0.813	0.563	0.688	0.313	0.750

Table 4. The mean, standard deviation and inter-quartile values for the three signals.

8.2.5 Conclusions

The evaluation of the results emerging from this subjective perceptual test may be considered positive. The objectives of the test have been fully achieved. It has been demonstrated that the technique developed allows a 3D binaural sound spatialization offering greater realism and quality as compared with four other techniques and algorithms corresponding to a significant sample of all of the binaural spatialization algorithms present both in the market and in the world of research (*see* Chapter 2). An

estimation of the standard deviation of the sample mean has been made. It shows that there is no overlap between the different score averages, at least not between the first engine and the other four. Furthermore, the distance in terms of standard deviations between the first engine and the others confirmed the robustness and significance of the results: for the comparison only between the first and the fifth engines, fewer than two standard deviations were found between the two means.

It has been proved that the process for the characterization of the BRIRs did not alter the localization cues within the impulses. Thus, the realism, quality and accuracy of a convolution-based spatializator and environmental simulator could be preserved, while retaining the flexibility of an algorithm-based option.

8.3 Possible future tests

The two subjective tests described within this chapter were considered to be satisfactory for providing a correct estimation of the quality of the various functions of the binaural spatialization algorithm.

Nevertheless, following on from the ideas and the results of these, two other tests could be carried out in the future, in order to verify the effectiveness of the algorithm itself when used in different applications.

A list of the suggested future tests follows.

- **Continuation of the comparative spatialization quality test**: the test described in Section 9.2 could be repeated using a larger number of binaural spatialization algorithms, e.g., including Ircam Spat 4.0. It was not possible to test this before because when the test was carried out, Spat 4.0 was only an unstable "beta" version. Also, the IEM AmbiTOBin PD (*see* Noisternig 2003a and 2003b, and Musil 2005) platform could be useful. It was not possible to test this before because when the test was carried out, no stable Universal Binary MacOSX version of the platform was available. The test could then be modified in order to be able to evaluate also localization accuracy, front-back confusion and distance perception.
- **Virtual-Real comparative spatialization quality test**: applying the same framework and platform of the test described in Section 9.2, instead of assessing the quality of different binaural spatialization algorithms and techniques, a test could be carried out comparing the binaural spatialization algorithm developed within this

research work and the simulation of real sound sources displacing loudspeakers around the subject. Wearing a pair of open headphones, a blindfolded subject could be presented alternatively with stimuli coming from the loudspeakers displaced within the room and from a virtual simulation of these loudspeakers presented through the headphones. Extensive research should be carried out in advance such that the headphones would not represent an obstacle for the signals incoming to the ear canal from sound sources within the surrounding environment. The test could then be modified in order to evaluate other factors such as localization accuracy, front-back confusion and distance perception.

- **Training session:** before every test, short training sessions could be performed, for example, giving to the subject tracking devices or simply a mouse and thus the opportunity to move a sound source in three dimensions. Also, presenting a series of sounds spatialized in different positions, played sequentially and accompanied by a visual representation of the rendered sound field, may broaden the options. Examples include having an image of the head in the centre and the sound source in the position where it is being simulated acoustically, or being seated in front of a wide screen where the sound sources appear graphically in the position where they are simulated acoustically. The tests could also be carried out twice, once before and once after the training session. An analysis of the data could indicate the responses to similar simulated auditory situations. Obviously, the test cannot be repeated in exactly the same form, although sufficiently informative similarities between the situations may be replicated.

8.4 Conclusions and global outcomes

Conclusions have been drawn individually for each of the perceptual tests performed. While the results of the first test were particularly useful in proceeding with the development of the binaural tool technique and application, the results of the second test were successful in verifying the effectiveness, quality and perceived realism of the final version of the binaural tool, and in particular of the environmental simulation technique. In Section 8.3, a list of possible future tests was offered. The results of this test are, nevertheless, considered sufficient in accomplishing the goals of the actual research work.

8.5 Brief summary

In this chapter the planning, the development, and the analysis of the results emerging from two perceptual subjective tests, carried out after the first and the last phases of the research work, have been described. The first test, detailed in Section 8.1, acted as the "work-in-progress" verification of the first version of the distance and environmental simulation techniques, and provided important feedback for the further development of the binaural tool. In contrast, the second test represented a global validation of the final version of the tool developed. It achieved highly important and positive results, confirming the spatialization quality and realism offered through using the techniques developed, and therefore that the goals established at the beginning of the research work have indeed been achieved.

In the last part of the chapter, a list and a description of possible further and future tests on the binaural tool were given.

Chapter 9

9. Possible Applications of the Developed Tool

Thus far, the thesis has focused on the design, developing, programming, and implementation of the binaural tool, including the previous chapter about the perceptual tests performed at the end of the research phase. It is nevertheless true that an important goal of the research is the application of that which has been developed: the more applicable the work, the more useful it might become, and the more successful the research will be. As already outlined (*see* Chapter 2), this is not the first binaural spatialization tool to have been created; other, similar, algorithms are already present on the consumer and professional markets. Five chapters (Chapters 3 to 7) have illustrated the aspects that make this tool particularly innovative and unique, as compared with others. While the model described is not complete (much can still be advanced, as will be summarised in Chapter 10), its strength and the new opportunities it offers could easily be used by, for example, artists, sound designers, and audio engineers.

In this last chapter before the conclusions, various applications of the tool will be discussed from the practical and the artistic points of view, suggesting possible uses for the application developed.

9.1 The Binaural Tool and multichannel live performances

When listening to a multichannel live performance, the listeners are all placed in the approximate middle of an array comprised of a given number of loudspeakers through which the audio performance is reproduced. In the exact centre of this loudspeaker system, within an area depending strictly on the speakers' configuration and on the technique used for the rendering of the surround audio scene (2D or 3D), there is the so-called 'sweet spot'. This is the area in which the soundfield is properly rendered, and outside of which variations from the original will be heard. The area of the sweet spot may be relatively small when compared with the total area in which the listeners are seated. It may, however, be enlarged by increasing the number of loudspeakers, the distances between the loudspeakers, and the use of particular 3D audio encoding and decoding techniques, such as High Order Ambisonics, VBAP or Wavefield Synthesis (*see* Chapter 2). Nevertheless, in a standard 2D eight-channel set-up, with loudspeakers

placed in the angles of an octagon with a diametre of eight metres, and in which panning between different speakers and therefore movements of the virtual sound sources are performed using only differences in levels, the sweet spot can be approximated as being a circle with a diametre smaller than one metre in the centre of the octagon. Obviously, in such a situation only a small segment of the audience would be able to sit in the sweet spot area, and the others would hear a modified soundfield.

In a discussion of this eight-loudspeaker 2D set-up and of the positioning of the audience in the centre of such a system, another phenomenon, already described in Section 3.6.3, gains high importance: the precedence effect. This effect, while being extremely useful for the localization of sound sources in a reverberant environment, may create problems when 3D (also 2D and 1D) soundfield simulations are performed using loudspeakers. If a listener is seated, for example, two metres closer to the two loudspeakers on the left, a stimulus coming from those loudspeakers will arrive six milliseconds before the same stimulus from the speakers placed on the right even if the signal played back through the system is exactly the same over the eight loudspeakers. This results in the listener's not hearing the virtual sound source as being central, but as being clearly located on the left, coming from the closer speakers. Considering only the precedence effect, which can be activated with delays greater than one millisecond between two speakers, it is obvious how even displacements smaller than one metre from the centre of the loudspeaker array can cause perceptible variations in the simulated soundfield. Of course, this effect is much stronger when exactly the same signal is reproduced from all of the loudspeakers simultaneously; furthermore, it is also true that if the listener is closer to one loudspeaker than to another, virtual sound source movements, sudden changes within the soundscape and, more generally, the whole sound scene may be altered perceptibly.

Given this precedence effect, together with the concept of the sweet spot, it may be stated that only a few members of an audience are able to perceive the recreated soundfield with a high level of accuracy, while variations would be present for all other members. The audience needs to be seated in an optimal position according to the shape and the location of the different loudspeakers, the positions of which dimensions are often very limited. As already stated, there are various 3D audio techniques aiming at enlarging the sweet spot area (*see* Section 2.4) maintaining the loudspeakers configura-

tion the same as the one of standard panning techniques; however, these are not widely used. It is due partly to their complexity in terms of encoding and decoding, and partly to the fact that no standards exist; widely used applications (such as Digidesign Pro Tools, or Apple Logic Audio) do not inherently support such formats and algorithms. Nevertheless, even were these applied, the majority of the audience would sit outside the sweet spot area, unless concerts are given only to a very small audience. In Kyriakakis (2002) an overview is given of several immersive audio techniques and of the technological limitations impeding the development of seamless immersive audio systems. One of Kyriakakis' proposed solutions uses a listener-tracking system, yet of course this creates limitations in terms of the number of possible simultaneous listeners (typically, one). Further research carried out at the ISVR in Southampton[1] (*see* Rose, 2002) investigated the dimensions of the sweet spot of virtual imaging acoustic systems at asymmetrical listener locations. In this case, too, the solution towards always having the listener inside the sweet spot implies the use of a tracking system, with the consequent limitations expressed above.

It is exactly in this scenario that the binaural tool could be used. Instead of having a complex loudspeaker system, possibly with multiple tracking devices, every listener could be equipped with a pair of headphones. These could be wireless, in order to minimize the problems generated by the presence of cables between the seats. The audience would thus listen to a properly recreated 3D soundfield, experiencing the exact multi-dimensional sound result planned by the composer/performer for recreation during the performance. Such a method would allow a far more closely controlled and controllable surround-sound experience. It would also provide the opportunity to create individual or group differences within the recreated sound scene, to change individually the listening volume, better to isolate the sound from environmental noises, and therefore expand the limits posed by any loudspeaker-based surround-sound system.

However, two issues might be raised at this point:

- The binaural simulation should be performed using individually measured HRTFs, in order to guarantee a high quality and efficient spatialization system;

1 *See* http://www.isvr.soton.ac.uk

- The presence of the headphones on the head of each listener might generate in the listener a sense of isolation from other members of the audience, rather than its being a shared experience, or cause other forms of discomfort during the performance.

Regarding the former issue, the work carried out within this PhD research suggests that it is indeed possible to create an accurate 3D binaural simulation without using individually measured HRTFs. With appropriate environmental simulations (as described in Chapters 5 and 6), the realism and quality of the 3D reproduced soundfield may increase such that individual HRTF measurements seem no longer to be essential. Similar to this problem is that linked with head-tracking, the fact that rotating the head while wearing headphones corresponds to a rotation of the simulated soundfield; it should normally remain unvaried and fixed within the space. A discussion of this aspect has already been given in the introduction to this PhD work.

Considering then the latter issue, the response is certainly more complex, and may depend on many factors; they, too, are linked with the characteristics of the performance and with the individuals comprising the audience. No solution can actually be reached without attempting such a performance and acquiring feedback, possibly through a questionnaire, then adapting the system in direct response to the comments and requests made by the audience.

It is nevertheless true that binaural spatialization, and in particular the binaural tool developed within this research, offers a particularly functional alternative to a standard loudspeaker system in multichannel sound performances, and that first steps should be taken towards the testing of these possible alternatives.

9.2 3D home audio binaural systems

Given that which has been stated in the previous section, a similar situation can be found when setting up a surround-sound system at home, in front of the TV, or for music listening purposes. Home audio surround systems have become much cheaper in the past ten to fifteen years. Nevertheless, in order to set up the optimum 2D or 3D audio system at home, many factors need to be taken into consideration: the positioning of all of the loudspeakers and the positioning of the listener(s); the acoustics of the room where the system is being installed, and any objects and obstacles located close to the speakers or to the listener which could block or alter the reproduced signals, for exam-

ple. It is most common to find home audio surround systems with loudspeakers positioned on bookshelves, partially obscured by books or other objects, pointing in directions other than towards the centre of the system, with a consequent decrease in the quality of the whole spatial audio experience. To such aspects need to be added all of the issues outlined in the previous section. Furthermore, in smaller environments and therefore with smaller distances between the loudspeakers, the sweet spot becomes smaller; the area where the listeners may be located in order optimally to perceive the simulated 3D soundfield may shrink to twenty to thirty centimetres in diametre.

It may easily be imagined how binaural audio could be useful within such a scenario: a 3D binaural audio encoder implemented directly inside a DVD audio player in order to be able play back 5.1 audio streams over a pair of headphones, would obviously offer a suitable solution. Other factors playing a role in this solution include the acoustics of the room where the sound is reproduced; the quantity of reflective and absorbent surfaces around the listener and the reproduction system, and their relative positioning. These would be utterly irrelevant to the optimum 3D surround sound perception. A headphone-based binaural system would be much less expensive than a 5.1 loudspeaker-based one. When listening to binaurally spatialized signals, no particular typology of headphones is required; the 3D surround sound effect would be present even when using cheaper makes.

Similar binaural systems have already been implemented and commercialized (*see* Dolby Headphones or SRS Headphones), although their quality is far from being satisfactory (*see* Chapter Two). The implementation within consumer DVD players or amplifiers of a complete and high quality binaural tool, such as that developed in the present thesis, would certainly offer a real alternative to loudspeaker-based surround sound systems.

A further application of the binaural tool is suggested for this particular home scenario. Many composers nowadays create multichannel musical compositions for reproduction through 2D or 3D loudspeaker arrays. Of course, these compositions are intended to be reproduced within concert halls, as outlined in the previous section; nevertheless, it often happens that a stereo mix-down is created for commercial purposes, such as CDs or DVDs. Within these stereo mix-down tracks, the spatial attributes of the original composition are reduced to left and right lateralization, with some depth gained thanks

to reverb simulations. Binaural spatialization could then be used for converting the multichannel tracks into a single 3D stereo track, preserving the spatial features and allowing the composition to be saved and diffused using standard media formats, such as CD-DA, for which only stereo tracks are allowed. Through approaching the situation in this way, merely a standard CD player and a pair of headphones would be needed in order to listen to and appreciate a full 3D audio soundscape.

Once the first version of the binaural tool was implemented (after February 2007), the author and different composers collaborated in order to attempt binaural conversions of multichannel surround musical pieces to be included within CD albums. Highly important feedback was gathered from these collaborations. It was acquired mainly from the composers, and also from other listeners, in order further to calibrate and ameliorate the whole spatialization algorithm and application. In particular, after each of the collaborations a survey was made regarding the quality, realism, and possible employment of the tool developed. The questions posed in the survey are reported here.

- How realistic was the simulation, and to what extent was it comparable with a multichannel loudspeaker listening?
- How much colouration was added to the sound by the binaural processing, and how acceptable was this compared to the colouration added by a loudspeaker system?
- What was the overall quality of the binaural processed signal, most of all if compared with other binaural spatializators used before (if applicable)?
- To what extent can binaural create, suggest or enforce different aesthetic approaches (or limitations) compared with other sound listening modes, such as mono, stereo and multichannel?

All of the work performed between the author and different composers and sound engineers during the whole Ph.D. research period (between January 2006 and August 2009), as well as relevant information given by each composer through the survey, is now detailed.

- **Leigh Landy** (composer, MTIRC, Leicester, UK; Professor Landy is also the First Supervisor of the author of this thesis), '*Oh là la radio*' (2006/7), commissioned by INA/GRM[2] (Paris, France) and published in the CD Bouquet of Sounds, MTIRC, MTI 001/2, 2007. Conversion between 2D eight-channel and binaural. The com-

2 *See* http://www.ina-entreprise.com/entreprise/activites/recherches-musicales/index.html

poser reported that the simulation was partially realistic: channels 1 and 2 seemed to be elevated and not completely frontal, but despite this the result was considered as being better in terms of spatial reconstruction accuracy than a possible stereo version (even if this was not the actual purpose of the composer). No particular sound colouration was heard on the spatialized signals, and "many people have found listening to the recording stimulating as they can hear the 'intimacy' I sought in concert halls of many sizes and shapes." Potential new opportunities provided by the binaural processor are seen for future applications, although in this case the goal was simply to convert a piece that already existed through performing an eight-channel binaural simulation: "The potential for new opportunities is there and might be worth pursuing in the future." As a final consideration, the composer stated:

> This is the best such musical recording system I'm aware of, having heard NASA surround recordings in the past. I look forward to the problem I encountered being resolved as it would act as an incentive to use it as a 'driver' for future recorded surround/3D work.

- **Peter Batchelor** (composer, MTIRC, Leicester, UK), *Kaleidoscope: Arcade* and *Kaleidoscope: Fissure* (2007), published in the CD Reflections, by Peter Batchelor, c3r records. Conversion between 2D twelve-channel and binaural. The composer reported that "spatialisation was excellent, and simulated the multichannel configuration very convincingly". Regarding sound colouration generated by the binaural processing:

  > For me it depended to some extent on the material. Certainly the binaural processing emphasised certain spectral characteristics, to which as composer I was inevitably hyper-sensitive. I preferred the outcome for *Kaleidoscope: Arcade* to that of *Kaleidoscope: Fissure* — I think because of the noisy character of the latter and reliance on high frequency content within this for detail, which was lessened slightly by the process. Inevitably there is a certain trade-off between the excellent spatial outcome and the slight colouration added. But the result overall was very compelling.

  No comparison could be made with a previously-used binaural algorithm, as this was the first time that the composer used such a technique; the result of the conversion performed by the algorithm was perceived as being "excellent".

- **Ron Herrema** (composer, MTIRC, Leicester, UK), *Let Freedom Ring*. Conversion between 2D eight-channel and binaural. Regarding the realism of the binaural simulation, the composer stated:

> Though the binaural experience of these works is certainly not functionally or aesthetically equivalent with the originals, I am mostly pleased with the outcome. As previews, the binaural mixes are generally superior to the stereo. The one exception to this is perhaps the opening of the concert version, in which the listener should experience a slow, circular envelopment of sound. The stereo technique I used for this — panning slowly from left to right — is, I think, a better surrogate than that which occurs in the binaural version. In much of the remainder of the music, however, which often involves a 'barrage' of point sources, the binaural rendition works quite effectively. In fact, one could argue that it works more effectively, since the experience is one of voices 'floating in the air' rather than emanating from a loudspeaker.

Regarding colouration brought by binaural processing: none was heard by the composer, who also stated that he was: "using low fidelity recordings and was not concerned with fine distinctions of colour." In his concluding comments on the new possibilities offered by the binaural technology applied to compositional processes, Ron Herrema stated:

> Finally, I think binaural sound suggests the possibility of a distinct, unique approach to electroacoustic composition. I do not think it is very useful for reinforcing ideas related to either multichannel or stereo composition, because these have nothing of the spatial realism of binaural sound. It is precisely the extreme realism of binaural sound that suggests new possibilities (and necessarily new constraints) in sound composition.

- Philippe-Aubert Gauthier (composer and sound engineer, GAUS, Sherbrooke, Canada), binaural conversions for the project *Frequences: Urbaines* of Karine Kotè, with the participations of various composers (Philippe-Aubert Gauthier, Robert Pelletier, Barah Hèon-Morisette and Guillaume Thibert). Conversion between various 2D and 3D multichannel formats and binaural. Referring to the binaural conversion of his pieces, the composer reported that "the binaural result was very convincing in creating a large extra-cranial sound space", while for other pieces the effect "was less convincing, especially for pieces that involved harmonic sounds with low attack rate." The composer also added:

> I had the chance to compare between the multichannel version and binaural version in a professional studio in 2009. I was then teaching

> spatial sound to sound artists, and I used my piece to illustrate the differences between multichannel and binaural versions of a composition; most listeners were very excited by the binaural version.

Regarding the colouration generated by the processing, Philippe-Aubert Gauthier reported that:

> Colouration was easy to control using the balance between the early reflections and direct sound. However, this was coupled to the resulting spatial impression. In all case, the colouration provided a more "natural" sound, with some very little added grain or texture.

Also that:

> The quality was excellent. The effect of not only using free-field HRTF was impressive. Indeed, the possible mix of HRTFs including early reflection and reverberation was a key point in achieving a pleasant result. Using the combination of these three set of HRTFs with varying degree had a great effect on the externalization. For my own piece, this was especially clear for broadband distorted (using amplifier and loudspeaker cabinet simulations) sounds. My comment is based on comparison with other HRTF databases.

On new possibilities offered by the application of the binaural algorithm to compositional processes, comparing them with standard stereo and surround techniques, the composer continued:

> The high-quality binaural version that I created with Lorenzo's filters […] helped to zoom on the spatial properties of the spatial composition. The binaural version made them more audible, more clear.

9.3 Teleconferencing and telepresence

Even if teleconferencing and telepresence cannot justifiably be considered artistic applications of the tool developed, the possible use of binaural spatialization techniques within teleconferencing and telepresence applications has nevertheless been considered sufficiently important for investigation in this chapter.

Section 3.6.2 provided information on the Cocktail Party Effect: when a listener is in a situation where more people are talking at the same level in one single room, s/he is nevertheless able to concentrate his/her attention on one single voice, isolating it from others even without rotating the head towards the speaker. This is possible thanks to the mechanisms of spatial hearing and Auditory Scene Analysis (*see* Bregman, 1990), which allow the hearing system to isolate different speakers located in different posi-

tions and to focus on one of them. It has also been established that if the listener uses a hand to close one of his/her ears, the ability to isolate and understand this single speaker becomes much weaker, and therefore the intelligibility of the speech decreases significantly.

In the context of teleconferencing applications, the voice of one or more participants is recorded with a microphone in one specific location, mixed with those coming from other places, and transmitted back to all of the locations at the same time. Usually, all operations are performed using monophonic audio files, and rarely with stereophonic. The recording of a complex soundscape with multiple competing speech signals through only one microphone, transmitting the signal to another location, then reproducing it back through a loudspeaker or a pair of headphones, results in complete erasure of all of the spatial attributes of the recorded soundscape. In this situation, the hearing system of a listener is far from having all of the information essential in performing the operations allowing the Cocktail Party Effect, i.e., the ability to focus on one specific speech without its being masked by others. The result is of a perceptible decrease in the intelligibility to the listener of the speech, and the consequent impossibility of focusing on an individual speech when more people are talking at the same time – a situation occurring frequently in standard teleconferences.

A perfect solution to such difficulties could be provided by the use of binaural applications both for the recording and for the reproduction of the different signals: the different signals could be recorded using a dummy head (or another multichannel surround microphone, such as Soundfield[3], converting then the signals to binaural), collected and collated from the different locations, then reproduced through a pair of headphones. The spatial characteristics of the different signals would be preserved through this method, resulting in an increase in the speech intelligibility for each respective participant. Relevant work in this area may be found in Pulkki (2007) and Jukka (2007)).

Two approaches may be employed for this particular application. The first, similar to that described in the previous paragraph, would imply recording the different signals using a dummy head or another multichannel surround microphone, with the signal then converted to binaural. In such a way, it would be important properly to position the

3 *See* http://www.soundfield.com

dummy head, or surround microphone, in respect of the speaker's or speakers' positions; having more than one speaker in the same position in each of the different recorded environments must be avoided. In this case, however, the spatial attributes of the different speech signals would be much less relevant to the Cocktail Party Effect. The two channel signals would be collected and collated, then sent to the different participants; the spatial source configurations of each environment would thus be preserved, with a consequent increase in the intelligibility of speech for all participants.

A second approach, more complex and certainly more flexible, would be to record each single speech source individually, with a monophonic microphone; the signals would then be processed in order to obtain virtual soundscapes with the different sources positioned binaurally in different positions. Each individual speech signal would therefore be processed with a binaural spatialization algorithm, creating a virtual sound source for each participant in his/her respective position. The processed signals would then be collated and transmitted to each participant, who would listen through a pair of headphones to a virtual binaural 3D soundscape. The binaural tool would be used as a means of increasing the intelligibility of a complex sound scene that has competing speech signals emanating from different positions. A similar approach, even if applied to auditory displays for aircraft and shuttle pilots, may be found in Begault (1996) and Brungart (1996).

It follows that a similar implementation could be performed without using the binaural technique, simply by employing a certain number of loudspeakers placed around each of the listeners. Even if such a solution were simpler from a computational point of view since the different sources would need simply to be panned over a certain number of loudspeakers and not to be binaurally spatialized, the problems linked with the number of signals to be transmitted and to the difficulties in the set-up and portability of the reproduction system would make this option irrefutably less convenient and practical.

9.4 Virtual reality and more

Another obvious utilization for the binaural tool would be within virtual reality applications. 3D video technologies are rapidly expanding, and their usage can now be commonly found in films and videogames, where 3D binaural audio technology could

easily be integrated. The simplicity, flexibility and effectiveness of binaural systems would be of significant advantage to an industry – virtual reality – that is growing quickly. Playing video games or watching DVDs while hearing 3D binaural audio from the point of view of any of the virtual characters or from an external perspective is just one example of what could be achieved through using the binaural technique creatively.

These and many other applications will doubtless be implemented. Thus far, though, discussions have focused on simulating spatial audio over headphones, on re-creating what in nature already exists: three-dimensional soundscapes. The mechanisms of spatial hearing have been investigated and analysed, and three localization cues characterized and simulated. These are the Interaural Level Differences (ILDs), the Interaural Time Differences (ITDs), and Direction-Dependent Filtering (DDF). Within the simulation of a real environment, these parameters would all be coherent with the position of the sound source. For example, for a sound source placed at 60° of azimuth, the sound would reach first the right ear then the left (ITDs). Furthermore, it would be more intense at the right ear (ILDs) than at the left, and the sound would be filtered depending on the particular resonances of the outer hearing systems for that specific sound source location.

It must, however, be asked what could happen if the three localization cues were incoherent with the real position of the sound source. Of course, this is impossible in nature, and equally so in a standard soundscape simulation, when loudspeakers are placed in a 3D space. Achieving such incoherence is not impossible in a system based on headphones, where the signals sent to the hearing system are much more controllable, thus the whole reproduction system results in being considerably more flexible.

In this case, the binaural spatialization technique could be useful not only to simulate a real 3D soundscape, but also to create new soundscapes, i.e., environments impossible to find in the real world.

This seems to be one of the astounding new options offered by binaural spatialization and, more generally, by computer science applied to sound. While it could indeed be considered inessential to simulate a feature that already exists in nature, it is particularly interesting to create a feature that as yet has no existence in the real world. A particular approach to this topic can be found in Picinali (2009).

9.5 Brief summary

In this chapter information has been given about possible artistic applications and uses of the binaural tool developed. In the first section (Section 9.1), the problem of the precedence effect in multichannel live performance has been addressed, focusing on how the binaural tool could be of help in solving it. A similar issue has been raised in Section 9.2 concerning home surround sound systems, again proposing binaural spatialization as a possible solution. The topic of the binaural conversion of multichannel surround musical compositions has therefore been raised and discussed, and a list of collaborations conducted between the author and different musical composer have been outlined, focusing on the benefits of the feedback gathered during this stage for the calibration and amelioration of the binaural algorithm and tool. Section 9.3 discussed the topic of telepresence in telecommunication applications, again focusing on the possible benefits in terms of speech intelligibility using binaural spatialization in critical teleconferencing applications. In the last section, before the final summary, a brief overview of possible applications of the binaural tool related to virtual reality applications has been offered, focusing also on the use of the binaural technique for the simulation of non-real soundscapes.

Chapter 10

10. Conclusions

This concluding chapter summarises briefly the whole thesis; the outcomes of the developmental and evaluation stages of this research are mentioned. Finally, in the third and fourth sections of the chapter, information is given about possible future improvements and additions to the research, from the development of new functions and/or applications to the setup and performing of further perceptual tests.

10.1 Summary

In this section, a brief summary of each chapter is listed, covering all of the topics that have been addressed, described, and expanded within the thesis.

Chapter 1: Basic Notions

This chapter introduced digital signal processing and concepts of acoustics and psychoacoustics, focusing on the specific topics related to the subjects involved in this research. Basic definitions were provided of certain terms of which knowledge is essential to understanding the rest of the thesis. After a univocal coordinate system for locating a sound source in a 3D space was established, an overview of the anatomy of the external hearing system was given, followed by an introduction to the elements of digital signal processing and filter design. Section 1.5 outlined and explained different representations of an audio signal, and an introduction to basic psychoacoustic principles was made.

Chapter 2: The State of the Art in the Field of Sound Spatialization

The focus of this chapter was on the extensive research that has been and continuing research that still is being carried out into the state of the art of sound spatialization during the Ph.D. The chapter attempted to justify the reasons for the current research; to convince the reader of its originality; to establish the theoretical framework and the methodological focus of the research itself, and to evaluate the products and the approaches of other companies and/or research centres.

- Section 2.1. Brief descriptions were given of the most significant companies in the surround sound field, their formats, and their respective specifications.

- Section 2.2. Given its focus on the consumer products market, the section provided descriptions of the most famous and familiar binaural and transaural systems available for home entertainment use.
- Section 2.3 provided an overview of multiple drivers surround sound headphones.
- Section 2.4. This section provided an overview of binaural and transaural techniques and systems aimed at the professional market, thus on those advanced software and hardware systems that have been made available for professional sound engineers in and researchers into the field of virtual surround sound and binaural spatialization.
- Section 2.5. This section was organized into one comprehensive table where nine systems for the consumer market and five for professional users were described and analysed in some detail. It included information, reported schematically, on a listening test performed by the author in an attempt to evaluate the effectiveness and realism of the spatialization accomplished by that specific software or device.
- Section 2.6. In this section, other state-of-the-art techniques and systems for 3D sound recording and reproduction currently under research in various centres around the world were listed, described, and evaluated.
- Section 2.7. Specific researchers, research groups, and projects working in the field of binaural spatialization have been described and grouped according to six different research topics: distance perception; HRTF measurement or simulation; HRIR interpolation techniques; HRTF quality testing; physical models of the human ear, head and auditory system, and spatial hearing and vision.
- Section 2.8. The Ph.D. research was placed into context, and the guidelines were elaborated, taking into consideration all of that which was accumulated from this preliminary research stage.

Chapter 3: Binaural Phenomena for the Perception of the Angle

In this chapter, the mechanisms of spatial hearing related to the perception of the angle of incidence have been described and analysed, starting from the interaural differences and concluding at the monaural cues. Sections 3.2 and 3.3 delineated the Interaural Level Differences (ILD) and Interaural Time Differences (ITD), while in Section 3.4 the topic moved to the Direction Dependent Filtering (DDF). The mechanisms for the auditory localization of sound sources placed in the three planes were then illustrated and discussed in Section 3.5, analyzing the accuracy of these phenomena for the angles

of both the azimuth and the elevation. Finally, Section 3.6 gave a brief overview of a selection of the most significant binaural effects. It is important to underline that within this chapter no innovations were presented; this was simply an introduction to the binaural phenomena concerned with the perception of the angle of incidence.

Chapter 4: Measurement of an HRIR Database

After the description of the mechanisms of spatial hearing (Chapter 3) and the illustration of the basic notions of the simulation of a linear and time-invariant system (Chapter 1), within Chapter 4 attention moved towards the simulation of three-dimensional soundscapes over headphones, which comprises the main topic of this research. Next, the topic proceeded to the measurement of the IR from a dummy head system, reporting information on the measuring technique and system, azimuth and elevation sampling, and IR processing and editing. Regarding this first part, no innovative techniques were presented, as it constituted merely a summary of techniques for IR measurement already introduced by other researchers.

In contrast, the latter part of the chapter presented innovative experiments for the measurement of the HRIR database. When the experiments contributing to this thesis were carried out, the use of the sweep technique was exploited only for architectural acoustics tasks, while the best-known and most widely used HRIR databases were measured using other techniques, such as the MLS (Gardner, 1994; Algazi, 2001). Since then, only one HRIR database has been released using the sweep technique (the IRCAM Listen project[1]), yet significant differences can be outlined between this and the methodology described within this research thesis.

Chapter 5: Binaural Phenomena for the Perception of Distance

Following the study into the binaural phenomena relevant to the perception of the angle of incidence of a given sound (Chapters 3 and 4), in Chapter 5 the topic moved to the perception of distance, i.e., how the human hearing system is able to determine the distance from the listener of a given sound source. As outlined in the chapter's first sections, the perception of distance is an extremely complex process involving numerous different parameters of the sound input into the hearing system. Many of these parameters, such as the intensity of the sound source and the reverberation generated by the reflections of the sound on the environment, may also vary independently of the

1 *See* http://recherche.ircam.fr/equipes/salles/listen

actual distance between the source and the listener, rendering their estimation especially complex.

The chapter started with a description of the mechanisms involved in the estimation of the distance of a sound source; the Inside the Head Locatedness effect was analysed, then three distance cues were outlined and described. In the last section, an experiment performed within the Ph.D. research on the ILD variations for close sound sources was described, and a brief analysis of the results provided.

It should be noted that within this and the subsequent chapter (Chapter 6) the main innovations of this research work have been described and analysed. Chapter 5 consisted mainly in a review of the literature on the perception of distance (except for Section 5.3, where an actual original study was described); Chapter 6 focused on the simulation of distance cues and on the creation of a binaural reverb algorithm, both of which may be recognised as the main innovations presented within this research.

Chapter 6: Distance Simulation and Binaural Reverb

In this chapter two intimately related innovative techniques have been presented and analysed. To simulate the distance of a virtual sound source, the anechoic source signal to be spatialized was processed in parallel with performing a convolution with three different HRIR and BRIR sets, i.e., the direct path signal, early reflections, and the reverberant, corresponding to the specific angles of azimuth and elevation of the position to be simulated. The convolution with the direct HRIR, created by isolating the direct component of the pseudo-anechoic HRIR, was made through performing a linear interpolation between the HRIR measured at different distances. The two BRIR sets were processed through the cross-synthesis algorithm in order to simulate different required acoustic environmental characteristics; next, a parallel convolution was performed with the source signal to be spatialized. The three generated spatialized signals were then combined, weighting the multiplication coefficient of each according to the distance to be simulated, and altering the pre-delay of the early reflections. Finally, the output signal was processed through a gain reduction line and a low-pass variable equalization filter, in order to simulate the frequency-dependent and frequency-independent components of the air absorption.

Chapter 7: The Binaural Spatialization Tool

In the context of the simulation of 3D soundfields over headphones, the previous chapters have presented and closely analysed innovative techniques for the simulation both of the positioning angle and of the distance of a given virtual sound source, as well as for acoustic environmental simulation. The majority of this was described in Chapters 4 and 6. In order to make the outcomes of this research project available to third-party users, in particular to musicians and composers, in this chapter the organization and implementation of the real-time and offline binaural processing software were analysed and described. The reason why two different items of software were created may be justified through two significant factors: firstly, the requirement in terms of CPU calculations of binaural processing with distance and environmental simulation is too large to allow a real-time version of the different modules, and therefore a real-time lighter implementation of the algorithm is necessary for monitoring purposes. Secondly, the offline binaural modules do not possess moving sound source functions; these thus need to be created in the Ambisonic domain, allowing real-time monitoring, using again a lighter implementation of the algorithm, and then converting the signals to binaural using the more complex offline modules.

Chapter 8: Subjective Perceptual Tests

In this chapter the planning, the development and the analysis of the results emerging from two perceptual subjective tests, carried out after the first and the last phases of the research work, were described. The first test, described in Section 8.1, acted as a work-in-progress verification of the first version of the distance and environmental simulation techniques, and gave important feedback for the further development of the binaural tool. In contrast, the second test represented a global validation of the final version of the developed tool; it demonstrated highly important and positive results, confirming both the spatialization quality and realism offered using the techniques developed, and that the goals established at the beginning of the research work have been reached. In the last part of the chapter, a list and a description of possible further (and future) tests on the binaural tool have been reported and briefly discussed.

Chapter 9: Possible Applications of the Developed Tool

This final chapter, the last before the actual conclusions, provides information on the possible artistic applications and usages of the developed binaural tool. Section 9.1

discusses the problem of the precedence effect in multichannel live performance, focusing on how the binaural tool could be of help in resolving it. A similar issue was raised in Section 9.2 concerning home surround sound systems, again proposing binaural spatialization as a possible solution. The topic of the binaural conversion of multichannel surround musical compositions has therefore been addressed and discussed; a list of collaborations conducted between the author and different music composers was outlined, focusing on the benefits of the feedback gathered during this stage for the calibration and amelioration of the binaural algorithm and tool. Section 9.3 dealt with the topic of telepresence in telecommunication applications, focusing again on the possible benefits in terms of speech intelligibility using binaural spatialization in critical teleconferencing applications. In the penultimate section before the final summary, a brief overview was given of possible applications of the binaural tool related to virtual reality applications; these focused also on the usage of the binaural technique for the simulation of non-real soundscapes.

10.2 Outcomes of the research work

As has already been outlined in the introductory chapter, this research performed work relevant to developments in different areas related to the field of binaural spatialization. The problems of the most known binaural spatialization algorithms present on the consumer and professional market, as well as the outcomes of the main research centres on the field (*see* Chapter 2), were outlined. The research proceeded towards the development and implementation of different techniques in order to make binaural spatialization more realistic, effective, and usable.

The outcomes of this research can be summarised in the following points:

- Measurement of an HRIR and BRIR database using the sinus-logaritmic sweep technique, and editing of the measured IR in order to obtain three different sets of responses: a pseudo-anechoic set, an early reflections set, and a reverberant set have been created (*see* Chapters 4 and 6).
- Development of a binaural spatialization technique with Distance Simulation, based on the individual simulation of the distance cues, and of Binaural Reverb, based on the weighted mix between the signals convolved with the different HRIR and BRIR sets (*see* Chapters 5 and 6).

- Development of a characterization process for modifying a BRIR set in order to simulate different environments with different characteristics in terms of frequency response and reverb time (*see* Chapter 6).
- Creation of a real-time and offline binaural spatialization application, implementing the techniques cited in the previous points, and including a set of multichannel- (and Ambisonics) to-binaural conversion tools (*see* Chapter 7).

Additionally to what is summarised in the previous points, within this research project a series of formal and informal perceptual tests was carried out on the applications and techniques developed. An initial perceptual test was conducted as soon as the first implementation of the distance simulation technique was available; the test confirmed the effectiveness of the theoretical background on the basis of which the binaural reverb and distance simulation technique were developed. This was simply a first step, allowing the improvement of the algorithm and its implementation within a more complete binaural spatialization tool.

A second, more thorough, perceptual test was performed on the final application (as described in Chapter 7), with the attempt to focus on the main innovation of the Ph.D. research: the spatialization of signals using a weighted mix of the different HRIR and characterized BRIR components. The test was based on a comparative task, aiming to evaluate the quality of the technique developed as compared with other environmental simulation techniques (stereo, multichannel, and binaural). The evaluation of the results may be considered extremely positive: the objectives of the test were fully achieved. It has been demonstrated that the technique developed allows a 3D binaural sound spatialization offering greater realism and quality when compared with four other techniques and algorithms; the latter correspond to a significant sample of all of the binaural spatialization algorithms present both in the market and in the research world (*see* Chapter 2).

To summarise the outcomes of the evaluation stage: while the results of the first test were particularly useful in order to proceed with the development of the binaural tool technique and application, the results of the second test represented a success for the verification of the effectiveness, quality, and perceived realism of the final version of the binaural tool, and in particular of the environmental simulation technique.

While this is simply a summary, the ways in which this research has contributed to the state of the art in the binaural spatialization field through introducing novel knowledge and tools have unequivocally been demonstrated. The contributions will, it is assumed, continue to be tested and further improved in the following years, in the attempt to approach a binaural simulation technique offering the same levels of realism and effectiveness as does a real-life 3D sound environment.

10.3 Potential future improvements

Despite the fact that the achievements outlined in the two previous sections have been considered as being sufficient for the fulfillment of a Ph.D. research degree, further improvements on the research so far developed may be foreseen. The author, possibly together with other researchers and research teams, will undoubtedly carry out further research into the topic of binaural spatialization, focusing both on areas of development and on the application of binaural techniques. A list of some of the potential improvements may be found below.

10.3.1 Further subjective testing on binaurally spatialized signals

As has been outlined at the end of Chapter 8, further perceptual tests could be carried out on binaurally spatialized signals. Additional studies could be performed on the estimation of the distance of a simulated sound source, in investigating, for example, the performances in environments with different acoustic characteristics (such as differing RT60s, or differing pre-delays for the early reflections components), and drawing comparisons between real and virtual environments.

Further studies could also be performed into the detection of other parameters linked with spatial hearing, such as the perception of different reverb typologies, the directional perception of early and late reflections, and the localization accuracy in different reverberant environments.

Moreover, experiments could be conducted into the respective performances of binaural perception before and after having carried out a “training session”, through giving to the subject, for example, the opportunity to “move” a sound source in the three dimensions (with tracking devices, or simply with a mouse). The auditory feedback could thus be linked with the visual, in order for the subject to adapt more rapidly to the binaural

simulation system. An example of studies of this nature may be found in Bronkhorst (1999).

10.3.2 Further studies into the splitting of the different BRIR components

The operations for the splitting of the different HRIR and BRIR components, for the characterization of the early reflections and reverberant IRs, for the calibration of the weighted sum between the signals spatialized with the different IR sets, and for the choice of the different air absorpion parameters (as described in Chapters 4 and 6) could be further expanded and improved, for example, in the attempt to measure longer BRIRs and the simulation of different environments with different acoustic characteristics (not only those related to the RT60 for the different frequency bands). The oucomes from this possible future research stage could introduce an increase in the flexibility, applicability, and effectiveness of the whole binaural tool.

10.3.3 Further studies into the variation of the ILDs for close sound sources

At the end of Chapter 5 (*see* Section 5.3), an investigation into the ILD variations for close sound sources was carried out, outlining interesting results for sound sources located between one metre and 20 centimetres from the head of the listener. Nevertheless, further development is required on this particular topic. Firstly, a closer analysis of the output data should be made; secondly, the development of a binaural simulation technique for close sound sources should be executed. The technique developed could then be included in the binaural tool, in order to offer to the user expanded functionalities related to the simulation of the distance of virtual sound sources.

10.4 Potential future additions

A list of some of the potential future improvements to the Ph.D. work has been created; nevertheless, in this section potential future additions, future opportunities for expanding this research, are listed and commented upon.

10.4.1 Perceptual studies concerning the introduction of incoherence between the three localization cues

At the end of Chapter 9 (*see* Section 9.4), a question was raised about the perceptual effect stemming from creating incoherence between the three localization cues within a

binaurally spatialized soundscape; for example, with a signal simulating the ITD of a sound source located at 60° of azimuth and 0° of elevation, the ILD of a source at 280° of azimuth and 0° of elevation, and the DDF of a source at 180° of azimuth and 90° of elevation. Similar investigations were already been carried out in Section 3.3, citing different research related to the Trading Difference and to the ITD vs. ILD competition. Perceptual studies on the area date, though, from the 1960s and 1970s; given the fact that the computational power of computers nowadays allows far more complex simulations to be tested, further development would surely be welcome, from both a psychoacoustic and an artistic point of view.

10.4.2 Binaural spatialization in audiology and audiometry applications

Various references have been made within the whole thesis to spatial hearing effects such as the Cocktail Party Effect and Binaural Masking (*see* Section 3.6). Such effects allow an improvement of the intelligibility of various signals, mainly speech, thanks to the spatial characteristics of the sounds input into the hearing system. In the context of surround soundfield reproductions, the simulation of a proper 2D or 3D virtual acoustic environment could therefore significantly influence the intelligibility of the signals reproduced (a similar approach was also used in Section 9.3).

This can undoubtedly be seen as an advantage in the development of algorithms to be implemented in hearing aid devices. In these devices, the sound is often picked up in non-optimal positions, such as behind the pinna (Behind The Ear or BTE hearing aids), and the spatial characteristics of the input signals are therefore distorted, resulting in a decrease in the intelligibility of speech in situations such as the Cocktail Party Effect. Further studies into this topic may also be found in Picinali (2008).

Further developments could also be made in the use of the binaural technique in the implementation of testing platforms for audiological examinations simulating real 3D listening environments, making the whole process more effective and realistic as compared with stereo simulations, and more flexible and applicable as compared with multichannel loudspeaker-based tests.

10.4.3 Binaural spatialization within VR applications for the blind

The question may be asked how it is possible for a blind individual to build a mental representation of a closed environment using only acoustic information. This could

constitute the starting point of a wide research project linked with binaural spatialization applied to Virtual Reality applications for the blind.

Nowadays, Virtual Reality applications are indeed mainly visually oriented; therefore, the majority of the information is transferred to the user through the visual channel. It is also a fact that the diffusion of multimodal interfaces is growing rapidly, allowing also non-sighted people to access Virtual Reality environments. Within this context, binaural spatialization techniques may easily be implemented within multimodal interaction devices in order to allow the exploitation of a 3D acoustic soundscape through a simple, inexpensive pair of headphones.

Another challenge could then be included within this possible framework: the performing of specific 3D audio psychoacoustic tests for non-sighted individuals. In fact, 3D audio algorithms are usually related to psychoacoustic principles based on perceptual tests carried out on sighted people. It is not too far from reality to assume that non-sighted individuals could show different performances in terms of spatial hearing as compared with sighted persons. An example of studies in this direction may be found in Doucet (2005).

Annotated Bibliography

[1] Algazi, V. R., Duda, R. O., Thompson, D. M., and Avedano, C. (2001). *The CIPIC HRTF Database.* Paper presented at the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New York, USA. Also: Gardner, B. & Martin, K. (1994), *HRTF Measurements of a KEMAR Dummy-Head Microphone*, MIT Media Lab Perceptual Computing, Technical Report #280, Boston, Massachussetts, USA.

These articles are reports of two of the best known and most widely used HRTF measurement databases: the CIPIC data were measured from different individuals, and the MIT data from a KEMAR dummy head.

[2] Blauert, J. (1996). *Spatial Hearing, the Psychophysic of Human Sound Localization*. Cambridge, Massachussetts, USA: The MIT Press Cambridge.

It can be considered the "bible" in terms of spatial hearing perception. Almost all of the experiments of the past fifty years in the spatial hearing field are reported within the book.

[3] Gerzon, M. (1974). Periphony: With-height sound reproduction. In *Journal of the Audio Engineering Society*, 21(1/2), pp. 2-10.

It is known to be the first article to mention "Ambisonics", thus is the first attempt to create a truly three-dimensional sound simulation technique. After this, hundreds of researches worked on the Ambisonics technique, expanding it from the first to the n^{th} order (High Order Ambisonics).

[4] Moore, Brian C. J. (2003). *An Introduction to the Psychology of Hearing*. London, UK: Academic Press.

To the same extent as the Blauert report may be considered the "bible" in terms of spatial hearing perception, this is at the same level in terms of the psychology of hearing and psychoacoustics.

[5] Pulkki, V. (1997). Virtual sound source positioning using vector based amplitude panning. In *Journal of the Audio Engineering Society*, 45(6), pp. 456-466.

The article in which Ville Pulkki introduces his VBAP (Vector Based Amplitude Panning) sound spatialization technique, definitely one of the most widely used in the past decade.

[6] Rayleigh, L (1907). On our perception of sound direction. In *Philosophical Magazine,* 13, 214-232.

Probably the first official publication about spatial hearing from an author who can be considered one of the fathers of modern acoustics.

Bibliography

[1] Algazi, V. R., Duda, R. O., Thompson, D. M. and Avedano, C. (2001). *The CIPIC HRTF Database.* Presented at the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New York, USA.

[2] Abbott, E. (1999, c1884). *Flatland: A romance of many dimensions*. London, UK: Penguin.

[3] Aschoff, V. (1963). Uber dar raumliche Horen (On spatial hearing). In *Arbeitsgem, D. ForshungNordrhein-Westfalen* 128, pp. 7-38, Westdeutscher Verlag, Koln.

[4] Batteau, D. W. (1967). The role of the pinna in human localization. *Proc. Roy Soc, London*, B168, pp. 158-180.

[5] Batteau, D. W. (1968). Listening with the naked ear. In S. J. Freedman (Ed.), *The neuropsychology of spatially oriented behavior* (pp. 109-133). Homewood, IL: Dorse Press.

[6] Begault, D. (1992). Perceptual Effects of Synthetic Reverberation on Three-Dimensional Audio Systems. In *Journal of the Audio Engineering Society,* Vol. 40, No. 11, pp. 895-904.

[7] Begault, D. (1996). *Virtual Acoustics, Aeronautics, and Communications*. Proceedings of the 101st AES Convention, November 1996, Los Angeles, California, USA.

[8] Begault, D., Wenzel, E. M., and Anderson, M. (2001). Direct Comparison of the Impact of Head Tracking, Reverberation, and Individualized Head-Related Transfer Functions on the Spatial Pereption of a Virtual Speech Source. In *Journal of the Audio Engineering Society*, Vol. 49, No.10, pp. 904-916.

[9] Begault, D. (2002). Challenging and Solutions for Realistic Room Simulation. In *Journal of the Acoustical Society of America,* Vol. 111, p. 2440.

[10] von Bekesy, G. (1938). Uber die Entstehung der Entfernungsempfindung beim Horen (On the origin of the sensation of distance in hearing). In *Akust. Z.* 3, pp. 21-31.

[11] von Bekesy, G. (1949). The moon illusion and similar auditory phenomena. In *Americal Journal of Psychology,* Vol. 62, pp. 540-552.

[12] Berkhout, A. J; de Vries, D., and Boone, M. M. (1980). A new method to aquire impulse responses in concert halls. In *Journal of the Acoustical Society of America*, 68(1), pp. 179-192.

[13] Blauert, J. (1996). *Spatial Hearing, the Psychophysic of Human Sound Localization*. Cambridge, Massachusetts, USA: The MIT Press Cambridge.

[14] Bosi, M. and Goldberg, R. L. (2003). *Introduction to Digital Audio Coding and Standards*. The Springer International Series in Engineering and Computer Science, USA.

[15] Bregman, A. S. (1990). *Auditory Scene Analysis*. Cambridge, Massachusetts, USA: The MIT Press Cambridge.

[16] Bronkhorst, A. W. (1999). Adapting head-related transfer functions to individual listeners (A). In *Journal of the Acoustical Society of America,* Vol. 105, No. 2, February 1999.

[17] Brookes, T. and Treble, C. (2005). *The effect of non-symmetrical left/right recording pinnae on the perceived externalization of binaural recordings*. Proceedings of the 118th Convention of the Audio Engineering Society, 28-31 May, Barcelona, Spain.

[18] Brungart, D. S., Durlach, N. I. and Rabinowits, W. M. (1999). Auditory Localization of Nearby Sources. In *Journal of the Acoustical Society of America*, Vol. 106, pp. 1465-1479.

[19] Brungart, D. S. and Simpson, B. D. (2001), *Distance-based speech segregation in near-field virtual audio displays*, Proceedings of the First International Conference on Auditory Display (ICAD 2001), Espoo, Finland.

[20] Burtgorf, W. (1961). Untersuchungen zue Wahrnehmbarkeit verzogerter Shallisignale [Investigations of the perceptibility of delayed sound signals]. *Acustica* No. 11, pp. 97-111.

[21] Capra, A. (2002). *Caratterizzazione quantitative e valutazione attraverso test soggettivi di teatri e sale da concerto [Quantitative characterizations and evaluation through subjective tests of theatres and concert halls]*. Graduate thesis for the "Laurea in Ingegneria Elettronica", supervised by Prof. A. Farina, Università degli Studi di Parma, Italy.

[22] Cherry, E. C. (1953). Some experiments on the recognition of speech with one and with two ears, *Journal of the Acoustical Society of America*, Vol. 25, pp. 976-979.

[23] Chowning, J. M. (1971). The simulation of moving sound sources. In *Journal of the Audio Engineering Society*, Vol. 19, pp. 2-6.

[24] Cook, Perry R. (Ed.) (1999). *Music, Cognition and Computerized Sound*. Cambridge, Massachussets: The MIT Press.

[25] Coleman, P. D. (1962). Failure to localize the source distance of an unfamiliar sound. In *Journal of the Acoustical Society of America*, Vol. 34, pp. 345-346.

[26] Coleman, P. D. (1963). An analysis of cues to auditory depth perception in free space. In *Psychological Bulletin*, Vol. 60, pp. 302-315.

[27] Coleman, P. D. (1968). Dual role of frequency spectrum in determination of auditory distance. In *Journal of the Acoustical Society of America*, Vol. 44, pp. 631-632.

[28] Doucet, M. E., Guillemot, J. P., Lassonde, M., Gagnè, J. P., Leclerc, C., and Lepore, F. (2005). Blind subjects process auditory spectral cues more efficiently than sighted individuals. In *Experimental Brain Research*, Vol. 160, No. 2, January 2005.

[29] Duda, R. O. and Martens, W. L. (1998). Range dependance of the response of a spherical head model. In *Journal of the Acoustical Society of America,* Vol. 104, No. 5, pp. 3048-3058.

[30] Durlach, N. I. and Colburn, H. S. (1978). Binaural phenomena. In *Handbook of Perception, Vol. IV*, edited by E. C. Charterette and M. P. Friedman, Academic Press, New York, USA.

[31] Farina, A. (2000). *Simultaneous Measurement of Impulse Response and Distortion with a Swept-sine Technique*. Presented at the 108th AES Convention, Paris, France.

[32] Farina, A., Glasgal, R., Armelloni, E., and Toger, A. (2001). *Ambiophonic Principles for the Recording and Reproduction of Surround Sound for Music*. Presented at the 19th AES International Conference, Shloss Elmau, Germany, 21-24 June.

[33] Farina, A. and Tronchin, L. (2005). Measurements and reproduction of spatial sound characteristics of auditoria. In *Acoustical Science and Technology*, 26(2), pp. 193-199.

[34] Feddersen, W. E., Sandel, T. T., Teas D. C. and Jeffress, L. A. (1957). Localization of high-frequency tones. In *Journal of the Acoustical Society of America*, Vol. 29, pp. 988-991.

[35] Fletcher, H. and Munson, W. A. (1933). Loudness, its definition, measurement and calculation. In *Journal of the Acoustical Society of America,* Vol. 5, pp. 82-108.

[36] Frannsen, N. V. (1960). *Some considerations of the mechanism of directional hearing.* Dissertation, Institute of Technology, Delft.

[37] Frova, A. (1999.). *Fisica nella Musica (Physics within Music)*. Bologna, Italy: Zanichelli.

[38] Fukuda, T., Horiuchi, T., Hokari, H. and Shimada, S. (2003). Relative distance perception by manipulating the ILD of HRTFs. In *Acoustical Science and Technology*, Vol. 24, No. 5, pp. 325-326.

[39] Gardner, M. B. (1969). Distance estimation of 0° or apparent 0°-oriented speech signals in anechoic space. In *Journal of the Acoustical Society of America*, Vol. 45, pp. 47-53.

[40] Gardner, B and Martin, K. (1994). *HRTF Measurements of a KEMAR Dummy-Head Microphone*. MIT Media Lab Perceptual Computing – Technical Report #280, Boston, Massachusetts, USA.

[41] Gerzon, M. (1974). Periphony: With-height sound reproduction. In *Journal of the Audio Engineering Society*, 21(1/2), pp. 2-10.

[42] Giesberts, T. (1995). SPDIF. Converting between AES/EBU and S/PDIF interfaces. In *Elektor Electronics Magazine*, July/August 1995, pp. 78-79.

[43] Grantham, D. W., Hornsby, B. W. Y. and Erpenback, E. A. (2003). Auditory spatial resolution in horizontal, vertical and diagonal planes. In *Journal of the Acoustical Society of America*, Vol. 114, No. 2, pp. 1009-1022.

[44] Greff, R. and Katz, B. FG (2008). Circumaural Transducer Array for Binaural Synthesis (A). *Journal of the Acoustical Society of America*, 123(5), pp. 3562-3562.

[45] Griesinger, D. (1996). *Beyond MLS – Occupied Hall Measurement with FFT Techniques*. Presented at the 101st Convention of the Audio Engineering Society, Los Angeles, USA.

[46] Hammershøi, D. (1995). *Binaural Technique: a Method of True 3D Sound Reproduction*. Ph.D. thesis, Aalborg Universitetsforlag, Denmark.

[47] Hansen, V. and Munch, G. (1991*).* Making Recordings for Simulation Tests in the Archimedes Project*.* In *Journal of the Audio Engineering Society,* Vol. 39, No. 10, pp. 768-774.

[48] Hartmann, W. M and Wittenberg, A. (1996). On the externalization of sound images. In *Journal of the Acoustical Society of America,* Vol. 99, No. 6, pp. 3678-3688.

[49] von Hornbostel, E. M. and Wertheimer, M. (1920). Über die Vahrnehmung der Schallrichtung [On the perception of the direction of sound]. *Sitzungsber. Akad. Wiss. Berlin*, pp. 388-396.

[50] Hwang, S. and Park, Y. (2006). *Time delay estimation from HRTFs and HRIRs.* Presented at the Eighth International Conference of Motion and Vibration Control (MOVIC 2006), Kaist, Daejeon, Korea.

[51] Jérôme, D. (2000). *"Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia",* Thèse de doctorat 1996-2000, Universitè Paris 6, Paris, France.

[52] Jukka, A., Pulkki, V., and Lokki, T. (2007). *Teleconference Application and B-Format Microphone Array for Directional Audio Coding*. Preprint of the Audio Engineering Society for the 30th International Conference, Saariselkä, Finland.

[53] Katz, F. G. Brian (1996). New approach for obtaining individualized head-related transfer functions. In *Journal of the Acoustical Society of America*, Vol. 100, p. 2609.

[54] Kietz, H. (1953). Das raumliche Horen (Spatial Hearing). In *Acustica*, Vol. 3, pp. 73-86.

[55] King, A. J., Kacelnik, O., Mrsic-Flogel, T. D., Schnupp, J. W. H., Parsons, C. H., and Moore, D. R. (2001). How Plastic Is Spatial Hearing. In *Audiology & Neurotology,* Vol. 6, No. 4, pp. 182-186,.

[56] Kirkeby, O., Nelson, P. A., and Hamada, H. (1997a). *The "Stereo Dipole" - Binaural Sound Reproduction Using Two Closely Spaced Loudspeakers*. Preprint of the Audio Engineering Society for the 102nd Convention, Munich, Germany.

[57] Kirkeby, O., Nelson, P. A., and Hamada, H. (1997b). *Virtual Source Imaging Using the "Stereo Dipole"*, Preprint of the Audio Engineering Society for the 103rd Convention, New York.

[58] Kistler, D. and Wightman, F. (1992). A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. In *Journal of the Acoustical Society of America*, Vol. 91, pp. 1637-1647.

[59] Kyriakakis, C. (2002). Fundamental and Technological Limitations of Immersive Audio Systems. In *Readings in Multimedia Computing and Networking*, edited by Kevin Jeffay and HongJiang Zhang, Academic Press, London, UK.

[60] Laws, P. (1971). *Sum Problem des Entfernungshorens und der Im-Kopf-Lokalisiertheit von Horereignissen* (*On the problem of distance hearing and the localization of auditory events inside the head*). Dissertation, Technische Hochschule, Aachen, Germany.

[61] Lithaud, A. (2003). *Audio Sculpt 2 user manual*. IRCAM, Paris, France.

[62] Mach R. (1865). Bemerkungen uber den Raumsinn des Ohres (Remarks on the spatial sense of the ear). In *Poggendorfs Ann.* 128, Fifth series, vol. 6, pp. 331-333.

[63] Malcangi, M. (2004). *Informatica Applicata al Suono [Computer Science Applied to Sound]*. Milano, Italy: Libreria CLUP.

[64] Malham, D. G. (1999). *Higher order Ambisonic systems for the spatialisation of sound.* Proceedings, ICMC99, Beijing, October.

[65] Malham, D. G. (2003). *Higher order Ambisonic systems*. Space in Music - Music in Space (M.Phil thesis). University of York. pp. 2–3.

[66] Martin, G. (2006). *Introduction to Sound Recording*. http://www.tonmeister.ca/main/textbook/ updated on 15/10/2006.

[67] Matsumoto, M., Tohyama, M., and Yanagawa, H. (2003). A Method of Interpolating Binaural Impulse Responses for Moving Sound Images. In *Acoustical Science and Technology*, Vol. 24, No. 5, pp. 284-292.

[68] McKeag, A. and McGrath, D. (1996). *Sound Field Format to Binaural Decoder with Head Tracking*. In Proceedings of the Audio Engineering Society for the Sixth Australian Regional Convention.

[69] Merimaa, J. and Pulkki, V. (2005) Spatial Impulse Response Rendering I: Analysis and Synthesis. In *Journal of the Audio Engineering Society*, vol 53, no. 12.

[70] Miller, J. D. (2001). *SLAB: A Software-Based Real-Time Virtual Acoustic Environment Rendering System*. Proceedings of the 2001 International Conference on Auditory Display, Espoo, Finland.

[71] Mills, A. W. (1958). On the minimum audible angle. In *Journal of the Acoustical Society of America,* Vol. 32, pp. 132.134.

[72] Mills, A. W. (1972). Auditory localization. In *Foundations of Modern Auditory Theory, Vol. II,* edited by J. V. Tobias, Academic Press, New York.

[73] Møller, H., Sørensen, M. F., Jensen, C. B. and Hammershøi, D. (1996). Binaural Technique: Do We Need Individual Recordings? *Journal of the Audio Engineering Society*, 44(6), 451/469.

[74] Moore, Brian C. J. (2003). *An Introduction to the Psychology of Hearing*. London, UK: Academic Press.

[75] Moorer, J. A. (1979). About this reverberation business. In *Computer Music Journal*, Vol. 3, No. 2, pp. 13-18.

[76] Muller, S. & Massarani, P. (2001). Transfer-Function Measurement with Sweeps. In *Journal of the Audio Engineering Society*, Vol. 49, No. 6, pp. 443-471.

[77] *Musil, T., Noisternig, M. and Hoeldrich, R. (2005).* A Library for Realtime 3D Binaural Sound Reproduction in Pure Data (PD). *Proceedings, Int. Conf. on Digital Audio Effects (DAFX-05), Madrid, Spain, 20-22 September.*

[78] *Nishimura, A. and Sasaki, M. (2004). Absolute cues for auditory distance in front and lateral directions. In* Acoustical Science and Technology*, Vol. 25, No. 2, pp. 127-135.*

[79] Noisternig, M., Musil, T., Sontacchi, A. and Hoeldrich, R. (2003a). *A 3D Ambisonic based Binaural Sound Reproduction System*, Proceedings, Int. Conf. Audio Eng. Soc., 26-28 June, Banff, Canada.

[80] Noisternig, M., Musil, T., Sontacchi, A. and Hoeldrich, R. (2003b). *3D Binaural Sound Reproduction using a Virtual Ambisonics Approach*. Proceedings, Int. Symp. on Virt. Env., Human-Computer Interf., and Meas. Sys. (VECIMS), 27-29 July, Lugano, Switzerland.

[81] Oppenheim, A. V. AND Shafer, R. W. (1975). *Digital Signal Processing*. Englewood Cliffs, N.J., USA: Prentice-Hall, Inc.

[82] Parks, T. N., Rubel, E. W., Popper, A. N., and Fay, R. R. (2004). *Plasticity of the Auditory System*. Springer, New York, USA.

[83] Paulo, J. P. and Coelho, J. L. B (2008). Swept Sine against MLS in room acoustics with music signals as background noise (A). In *Journal of the Acoustical Society of America*, Vol. 123, No. 5, pp. 3617-3617.

[84] Perrot D. R. and Saberi K. (1990). Minimum audible angle for sources varying in both elevation and azimuth. In *Journal of the Acoustical Society of America,* Vol. 87, No. 4, pp. 1728-1731.

[85] Picinali, L. (2006). *Techniques for the extraction of the impulse response of a linear and time-invariant system.* Presented at the DMRN Doctoral Research Conference, 15-16 July, University of London, UK.

[86] Picinali, L. (2007a). *The simulation of the distance in a binaural spatialization algorithm.* Presented at the first SPAce-Net Workshop, 25 January 2007, York, UK.

[87] Picinali, L. (2007b). *La simulazione della distanza in un algoritmo per la spazializzazione binaurale (The simulation of the distance in a binaural spatializazion algorithm)*. Presented at the 34th National Conference of the Associazione Italiana di Acustica (Italian Society of Acoustics), 13-15 June 2007, Firenze, Italy.

[88] Picinali, L., Prosser, S., Mancuso, A., and Vercellesi, G. (2008). *Speech Intelligibility in Virtual Environments Simulating an Asymmetric Directional Microphone Configuration*. Presented at the Acoustics '08 International Conference, June-July 2008, Paris, France.

[89] Picinali, L. (2009). 3D Sound Simulation over Headphones. In *Handbook of Research on Computational Arts and Creative Informatics,* edited by Braman, J.. Vincenti, G., and Trajkovski, G., Information Science Reference, Hershey, New York, USA.

[90] Pierce, A. H. (1901). *Studies in auditory and visual space perception, Vol. 1: The localization of sound.* Longmans, Green, New York, USA.

[91] Pulkki, V. (1997). Virtual sound source positioning using vector based amplitude panning. In *Journal of the Audio Engineering Society*, 45(6), pp. 456-466.

[92] Pulkki, V. and Merimaa, J (2006). Spatial Impulse Response Rendering II: Reproduction of Diffuse Sound and Listening Tests. In *Journal of the Audio Engineering Society*, Vol. 54, No. 1.

[93] Pulkki, V. (2007). Spatial Sound Reproduction with Directive Audio Coding. In *Journal of the Audio Engineering Society*, Vol. 55, No. 6, pp. 503-516.

[94] Rabiner, L. R. and Gold, B. (1975). *Theory and Aplication of Digital Signal Processing*. Englewood Cliffs, New Jersey, USA: Prentice Hall, Inc.

[95] Rayleigh, L (1907). On our perception of sound direction. In *Philosophical Magazine,* 13, pp. 214-232

[96] Reichardt, W. and Haustein, B. G. (1968). Zur Ursache des Effektes der "Im-Kopf-Lokalisation" (On the cause of the inside-the-head locatedness effect). In *Hochfrequenztech. U. Elektroakustik,* Vol. 77, pp. 183-189.

[97] Rife, D. D. and Vanderkooy, J. (1989). Transfer-Function Measurement with Maximum Length Sequences. In *Journal of the Audio Engineering Society,* Vol. 37, n. 6, pp. 419-444.

[98] Rose, J., Nelson, P., Rafaely, B. and Takeuchi, T. (2002). Sweet spot size of virtual acoustic imaging systems at asymmetric listener locations. In *Journal of the Acoustical Society of America*, Vol. 112, No. 5, pp. 1992-2002.

[99] Rosenlicht, M. (1985). *Introduction to Analysis*, Dover Publications, New York, US.

[100] Seraphim, H. P. (1961). Ueber die Wahrnehmbarkeit mehrerer Rueckwuerfe von Sprachshall [On the perceptibility of multiple reflections of speech sounds]. In *Acustica* No. 11, pp. 80-91.

[101] Seraphim, H. P. (1963). Raumakustische Nachbildungen mit elektroakustischen Hilfsmitteln [Simulation of room acoustics using electronic aids]. In *Acustica* No. 13, pp. 75-85.

[102] Schirmer, W. (1966). Zur Deutung der Ubertragungsfehler bei kopfbezueglicher Stereophonie (On the explanation of errors in hear-related stereophonic reproduction). In *Acustica*, Vol. 17, pp. 228-233.

[103] Schroeder, M. R. (1962). Natural Sounding Artificial Reverberation. In *Journal of the Audio Engineering Society*, Vol. 10, p. 219.

[104] Shaw, E. A. G. and Teranishi, R (1968). Sound pressure generated in an external-ear replica and real human ears by a nearby sound source. In *Journal of the Acoustical Society of America*, Vol. 44, pp. 240-249.

[105] Shutt, C. E. (1898). *Experiments in judging the distance of sound.* Kansas University Quart. 7, pp. 9-16.

[106] Smith, J. O. III (2009). *Physical Audio Signal Processing*. CCRMA Publications at Stanford, California, USA.

[107] Sone, T., Ebata, M. and Tadamoto, N. (1968). *On the difference between localization and lateralization.* Presented at the Sixth International Congress of Acoustics, 24-28 July, Tokyo, Japan.

[108] Sontacchi, A., Majdak, P., Noisternig, M. and Holdrich, R. (2002). *Subjective Validation of Perception Properties in Binaural Sound Reproduction Systems*. Presented at the 21st Conference of the Audio Engineering Society, 1-3 June, St. Petersburg, Russia.

[109] Svensson, P. and Nielsen, J. L. (1999). Errors in MLS Measurement Caused by Time Variance in Acoustic Systems. In *Journal of the Audio Engineering Society*, Vol. 47, No. 11, pp. 907-927.

[110] Takekuchi, T., Nelson, P. A., Kirkeby, O., and Hamada, H. (1998). *Influence of Individual Head Related Transfer Function on the Performance of Virtual Acoustic Imaging Systems.* Proceedings of the 104th Convention of the Audio Engineering Society, 16-19 May, Amsterdam, The Netherlands.

[111] Tokuno, H., Hamada, H., Kirkeby, O., and Nelson, P. (1996). Binaural sound reproduction in a stereo dipole system. In *Journal of the Acoustical Society of America*, 100, p. 2700.

[112] *Theile, G. (2004).* Wave Field Synthesis - A Promising Spatial Audio Rendering Concept. *Proceedings of the Int. Conf. on Digital Audio Effects (DAFX-04), 5-8 October, Naples, Italy.*

[113] *Thomas, M. V. (1977). Improving the Stereo Headphone Sound Image. In* Journal of the Audio Engineering Society, *Vol. 25, No. 7/8, pp. 474-478.*

[114] *Wall, K. and Von Hagen, W. (2004).* Definitive Guide to GCC (Expert's Voice). *APress, USA.*

[115] *Wallach, H., Newman, E. B., and Rosenzweig, M. R. (1949). The precedence effect in sound localization.* In *J. Exp. Psychol.*, *Vol. 27, pp. 339-368.*

[116] *Weinrich, S. G. (1992).* Improved Externalization and Frontal Perception of Headphone Signals. *Proceedings of the 92nd Convention of the Audio Engineering Society, 24-27 March, Vienna, Austria.*

[117] Yost, W. A. (2000). *Fundamentals of Hearing: An Introduction*. Academic Press, San Diego, California, USA.

[118] Zahorik, P. (2002). Assessing auditory distance perception using virtual acoustics. In *Journal of the Acoustical Society of America,* Vol. 111, No. 4, pp. 1832-1846.

Appendix A

Table of the Research Groups in the "binaural spatialization world"

This section has its own bibliography, separately from that of the PhD thesis. The reason for this is that the bibliography of this section is related only to the articles, books, and conference papers cited within the table (right-hand column); these have not always been read by the author, and may not be directly related to the topics discussed in the thesis. Any article, book or conference paper may be present both in this bibliography and in the general bibliography.
In the right-hand column, a brief bibliography of the most recent or most important publication of the different research groups is reported.
Most of the information within this table has been gathered from the respective Internet sites of the research centres.

NAME (internet site and date of latest information retrieved)	**TOPIC**	**RESEARCH FIELDS AND ACHIEVEMENTS**	**NOTES and BIBLIOGRAPHY**
The Acoustics Laboratory Aalborg University http://acoustics.aau.dk/ *Jan 2009*	The Department of Acoustics at Aalborg University is placed within the Institute of Electronic Systems, which means that our students usually have a background in electrical engineering. This means that the education is strong in the field of electroacoustics (loudspeakers, microphones and so on), but it does not mean that they only work in electroacoustics. Students at this department will also work with physical acoustics, room acoustics, building acoustics, psychoacoustics,	• **Distance perception:** How does human hearing judge distance to the sound source? In this project a large body of experiments was conducted in order to investigate how well human hearing judges distance. The main result was that reflections from the surroundings play a dominant role in distance perception. • **Sound transmission to and within the ear canal:** A theoretical and experimental analysis of how sound is transmitted to and within the human ear canal. A sound transmission model is made and verified by measurements. This model is used for improving binaural recording techniques.	**Henrik Moller, Wolfgang Ellermeier, Dorte Hammershoi, Karin Zimmer, Sofus Birkedal Nielsen, Flemming Christensen, Ville P. Sivonen, Pauli Minnaar, Pablo Faundez Hoffmann** and **Steffen Pedersen** work in this research group. BIBLIOGRAPHY (just some of the…): *For more information, look at web page* http://acoustics.aau.dk/publications/pubframe.html *Sivonen (2006), Hoffmann (2006),*

	psychometry, acoustical measurement tecniques, noise etc.	• **A loudspeaker with controlled directivity:** A loudspeaker box usually contains several loudspeakers because a large diaphragm area is preferable at low frequencies, while a small area is best at high frequencies. At low frequencies a large body of air has to be moved, which calls for a large diaphragm area, but large diaphragms tend to break up at high frequencies. The aim of this project was to construct a loudspeaker with controlled breakup and directivity by experimenting with different membrane geometries. • **Artificial head techniques for noise evaluation:** Traditional noise measurements determine sound pressure level or sound intensity using ordinary microphones. These measurements fail to take into account the directional dependent amplifications of the human head, and may thus be very inaccurate as measures of noise exposure. In this project artificial head techniques are investigated and meaningful assessment methods for the comparison of noise sources using artificial heads are developed. • **Designing an optimal artificial head:** The quality of the existing artificial (dummy) heads is evaluated objectively and subjectively; errors and imperfections are identified and an alternative approach is developed. The goal is to design a scientifically optimal dummy head for use in binaural recording and noise evaluation. • **Binaural auralization: Generating correct impressions of sound sources in rooms:** Know-	*Minnaar (2006), Hoffmann (2008a and 2008b),*

		ledge of sound transmission to and within the ear canal is used for artificial generation of binaural signals through digital signal processing. The purpose is to make it possible to listen to, e.g., a concert hall before it is built, or to loudspeaker systems during the development process. • **SCATIS: Spatially Coordinated Auditory/Tactile Interactive Scenario:** Existing Virtual Reality (VR) systems usually focus on the visual sense, and more or less neglect the other senses. In this project the auditory and tactile senses received full attention. A laboratory system for research into the interaction between touch and hearing was developed. • **Methods for designing and measuring headphones:** How does the perfect headphone sound? This is not a trivial question. In this project different design goals for headphones are developed, as well as reliable measurement methods. • **Multiprocessor DSP system for digital audio:** Many signal processing algorithms for digital audio can be implemented with advantage in a parallel structure. A multiprocessor DSP system with standard audio interfaces is developed, together with a debugger/monitor system allowing debugging on all processors simultaneously. • **Veridical sound for virtual reality systems:** The sound in existing virtual reality systems is not very convincing. The possibilities for improvement using binaural auralization are investigated. Virtual reality sound has several applications besides VR	

		systems, e.g., cockpit communication systems, teleconferencing and man/machine interfaces for process control and robot control. • **Methods for testing human hearing:** Methods for determination of hearing thresholds - monaural as well as binaural - are investigated, and hearing threshold data are gathered. The new methods are compared to existing calibration methods for headphones for the measurement of clinical hearing thresholds and are evaluated by a clinical example: "Does the use of Walkmen damage hearing?"	

Acoustics Research Centre
School Computing Science and Engineering
University of Salford
http://www.acoustics.salford.ac.uk/research/
 	
Sept 2007
	Salford University has a 30-year history of research into acoustics. Principle areas include building and architectural acoustics, environmental acoustics, outdoor sound propagation, remote acoustic sensing of metrological conditions, subjective response especially for room acoustics and audio systems, digital signal processing, transducer design and active control.	• **Development and Perception of Spatial Sound Reproduction Systems:** This project aims to investigate perceptual mechanisms in operation when people listen to spatially extended sounds reproduced by multi-loudspeaker systems. The information will be used to improve designs for such systems.
		  • Measurement system for comparing spatial performance of different reproduction systems completed
		  • Spaciousness of distributed harmonic sounds investigated
		  • Effects of reproduction systems on source localisation models investigated
		• **Clean audio project - surround sound quality**
		• **Processing for Binaural and Multi-Media Communications:** This research has investigated an objective measure of communication quality in multi-media systems. In addition it has developed a compact and hierarchical description of the bi-
			James Angus, Prof Stuart Bradley (visiting), Prof Trevor Cox and **Prof Yui Wai Lam** work on this research project.
			
			BIBLIOGRAPHY:
			Hirst (2000).

		naural sound field using spherical harmonics. Two other colleagues and I are now investigating the use of acoustic modelling techniques to derive person-alised head related transfer functions from photogrammetric data. This will allow a higher quality of sound spatialization in multimedia systems.	
AES Technical Committee on Multichannel and Binaural Audio Technologies http://www.aes.org/technical/mbat/ *Sept 2007*	This committee addresses fundamental production and reproduction issues of multichannel audio systems employing loudspeakers, as well as binaural techniques for creating multidirectional illusions through headphones and loudspeakers.	• **Loudspeaker Production / Reproduction Issues** • **Headphone Production / Reproduction Issues** • **Loudspeaker / Headphone Reproduction Conversion Issues** • **Head Related Transfer Functions** • **Headphone Externalization Techniques**	THOSE INVOLVED (some of): **Francis Rumsey (Chair), Gunther Theile (Vice Chair), Akira Fukada, Durand Begault, James Johnston, Jan Berg, Jyri Huopaniemi, Marina Bosi, Mick Sawaguchi, Nick Zacharov, Renato Pellegrini, Takeo Yamamoto, Theresa Leonard, Geoff Martin, Russell Mason, Ville Pulkki, Franz Zotter.** RECENT ACTIVITIES: • 19th International Conference on 'Surround Sound: Techniques, Tech-nology and Perception', 21-24 Junee 2001, Germany. Click here to order proceedings. • AES 24th International Conference on 'Multichannel Audio - The New Reality'. 26-28 Junee 2003. Banff, Canada. June • AES 28th International Conference on 'Future of Audio Technology - Surround and beyond'. 30 Junee-2

			Julyy 2006. Pitea, Sweden.
Spatial Auditory Display Laboratory NASA Ames Research Center. http://vision.arc.nasa.gov/PPSF/1-Perceptual/sub1-6/sub1-6.html *Jan 2009*	Develop auditory displays that prioritize and spatially segregate auditory information for improved situational awareness, intelligibility, and for reduced workload. Model potential and existing auditory environments in aviation contexts such as the flight deck. Combining 3-D audio technologies with active noise cancellation, the auditory display system controllers implement spatial audio technologues: separate channels of auditory information are placed at different virtual locations to provide situational awareness (e.g., airborne or ground traffic collision avoidance alerts; taxiway navigation aids and announcements), increase intelligibility (through the use of binaural delivery systems) and reduce auditory fatigue.	• **HRTF simulation:** The ASAD (Ames Spatial Auditory Display) simulation studies are supplemented by basic research in human sound localization and communications intelligibility. Validation of acoustic measurement and modelling techniques for reverberant environments are developed, validated, and refined, beginning with simplified models, and then increasing in complexity to achieve accurate modelling of the flight deck. Specialized hardware development and psychoacoustic validation of HRTF measurement and rendering techniques are required to enable these goals. • **SLAB:** It is a real-time virtual acoustic environment rendering system which performs spatial 3D-sound processing allowing the arbitrary placement of sound sources in auditory space. • **Snapshot measurement system**: Low-cost portable system for measuring HRTFs (Crystal River Engineering). • **HRTF measurements:** Recorded with Snapshot measurement system, using Golay code sequences. By constraining the geometry to ensure that no early reflections arrive within the first few milliseconds after the direct path arrival, the need for an anechoic chamber is eliminated -- simple windowing techniques are used to extract the desired HRTF from the measured impulse response. • **Simplified HRTF filtering**: Adapted for finite impulse response filtering on Motorola 56001	ASAD is the first 3D audio processor designed especially for multiple communication channels: it can place five different communication channels in virtual auditory positions about the listener by filtering each input channel with binaural head-related transfer function data. **Durand R. Begault** and **Elizabeth M. Wenzel** work in this research group. BIBLIOGRAPHY: *Wightman (1989; 1995), Wenzel (1992; 2003), Abel (1994; 1995), Begault (2005)*

		digital signal processors. • **Speech intellegibility for 3D audio.**	
The AUDIS catalogue of human HRTFs http://www.eaa-fenes-tra.org/Products/Documenta/Publications/09-de2 *July 2006*	In the European-Union-funded project AUDIS (AUditory DISplay), a multipurpose auditory display for 3D hearing applications is being developed. The main goal is to develop a special sound generator in combination with an auditory-symbol database that can, for example, be used in the avionic and automotive industries.	• **HRTF sets of 12 subjects** • **HRTF set of the dummy head HMS II** • **Tools for the HRIR Interpolation** • **Grafical representation of HRIR and HRTF** • **Convolution for monochannel** • **Demonstration of spatialized audio**	**Jens Blauert** and **Mark Brueggen (**Ruhr Univ., D-44780 Bochum, Germany**), Adelbert W. Bronkhorst** and **Rob Drullman** (TNO NL-3769 ZG Soesterberg), **Gerard Reynaud** and **Lionel Pellieux** (Sextant Avionique, F-33166 Saint-Medard-en-Jalles), **Winfried Krebber** and **Roland Sottek** (Head Acoust. GmbH, D-52134, Herzogenrath, Germany) work in this research group.
Auditory Perception Lab U.C. Berkeley, Dept. of Psychology http://ear.berkeley.edu/auditory_lab/ *July 2006*	The research is concerned primarily with issues involved in the higher-order processing of auditory information, especially as it impacts on listening in real world situations. The first half of our research programmeme concerns the localization of sounds in space; the second half involves the various types of attentional mechanisms available to listeners and the tasks that are associated with them.	• **Spatial Hearing with Complex Stimuli**: Studies into Interaural Time Differences, Echo Suppression and Precedence Effect • **Divided Attention, Task Demands and the Detection of Amplitude Fluctuations**: Where do we direct our attention if two independent stimuli are reproduced simultaneously at the two ears? • **Simulated Open Field Environment (SOFE):** Complex loudspeaker system (in an anechoic chamber) for the simulation of an open-field environment (it solves some of the problems of HRTF simulation).	At this link (**http://ear.berkeley.edu/**) plenty of information about the Hearing Sciences. There is also a picture database of external ear shapes. **Hervin Hafter** and **Anne-Marie Bonnel** and **Michael Pukish** work in this research group. BIBLIOGRAPHY: *Two chapters in Gilkey (1997). Hafter (1997)*
The Auralization and Acoustic Laboratory Naval Postgraduate School http://www.icad.org/websiteV2.0/Conferences/ICAD96/proc96/storms.htm	As an expansion of the NPSNET Research Group (NRG), the Auralization and Acoustics Laboratory (AA-Lab) at the Naval Postgraduate School studies the integration of aural cues	Research within the AA-Lab focuses on the ability of low-cost commercial sound equipment, with Musical Instrument Digital Interface (MIDI) (International MIDI Association, 1983), to produce aural cues for the distributed virtual environment of the Naval	**Russell Storms, Lloyd Biggs, William Cockayne, Paul Barham, John Falby, Don Brutzman** and **Michael Zyda** are part of this research group.

July 2006	into virtual environments. Currently, the AA-Lab focuses on spatial-acoustic sound rendering *via* headphones (closed-field) and loudspeakers (open-field).	Postgraduate School Networked Vehicle Simulator (NPSNET) (Zyda *et al*., 1993a, 1993b, 1995; Macedonia *et al*., 1995). This research specifically examines the ability of both headphone (closed-field) and loudspeaker (open-field) delivery systems to integrate aural cues in large-scale, distributed virtual environments that comply with Distributed Interactive Simulation (DIS). • **Headphone systems:** They are investigating alternative headphone delivery systems for NPSNET. Proposed headphone delivery systems must render spatially a minimum of eight simultaneous sound events in real time. One system produces spatial sound on the same workstation that renders the graphics of a virtual simulation. Another approach dedicates a workstation as a sound server, rendering spatial sound for multiple clients, which are connected via a local-area network. A third approach creates a library of prerecorded positioned sound files. This system requires the host CPU only to determine position information, to locate the appropriate sound file, and to execute the prerecorded spatial sound. • **Loudspeaker systems:** The NPSNET-3D Sound Server (NPSNET-3DSS) is a MIDI-based loudspeaker sound system consisting of commercial sound equipment and student-written computer software, software designed to generate 3-D aural cues via a cube configuration of eight loudspeakers and known as The NPSNET Sound Cube. Using an algorithm similar to stereo panning, the system	**About NPSNET:** The NPSNET Visual Simulation System is a real-time, interactive distributed simulation system that was developed by students at the Naval Postgraduate School (NPS). The system is written in C++ and uses SGI's Performer. NPSNET reads MultiGen Flight databases and is DIS-compliant. The system can also read in SIMNET models. NPSNET includes expert systems that control autonomous forces. Stereoscopic views can be generated by the software and displayed using StereoGraphics' CrystalEyes system. A modified version of NPSNET has been used in a dismounted infantry capability and walk-in synthetic environment demonstration at the 1994 Individual Combatant Modelling and Simulation Symposium, held at Fort Benning. The project involved NPS, the University of Utah, Sarcos and the University of Pennsylvania. BIBLIOGRAPHY: *Cockaine (1996), Macedonia (1995).*

		sends sound to the various speakers of the sound cube to create an apparent (phantom) image of the sound event relative to the azimuth and elevation of the listener. To enhance distance perception, the system adds synthetic reverberation, which is produced by two Ensoniq DP/4 Parallel Effects Processors, to discrete sound events via real-time MIDI modulation messages.	
Binaural Hearing Lab University of Boston http://www.bu.edu/dbin/binaural/?Section=links *Jan 2009*	The Binaural Hearing Laboratory is focused on studies of binaural interaction, including phenomena such as sound localization for which monaural processing also plays a major role. The goal of these studies is an integrated understanding of binaural interaction and its role in human sound perception including the interpretation of acoustic cues in complex sound environments (e.g., multiple sources in reverberant spaces).	Specific projects range from signal processing models of physiological activity to empirical measurements of the hearing abilities of listeners with hearing losses and/or neurological lesions. • **Basic Binaural Sensitivity**: The goal is to measure psychophysical performance under a variety of binaural conditions. It is focused on localization and masked detection tasks with perceptually simple stimuli as well as speech intelligibility under conditions with minimal uncertainty. • **Models of Brainstem and Midbrain Neurons:** Developing of computational models for the activity of neurons in the auditory brainstem and midbrain. The focus is on the binaurally sensitive neurons that respond to interaural time difference (ITD) and/or interaural level difference (ILD). The model simulations are based upon specific hypotheses on the effects of interactions between synaptic excitation and inhibition and of characteristics of membrane properties on ITD and/or ILD sensitivity of single binaural neuron in response to pure tones and amplitude-modulated tones. • **Performance of Bilateral Cochlear Implants:**	**H. Steven Colburn**, **Nathaniel I. Durlach** and **Jacob W. Scarpaci** work in this research group. BIBLIOGRAPHY: *Scarpaci (2004, 2005a, 2005b and 2006)*

		Measuring of the hearing abilities of bilateral CI users in several types of experiments, including sound localization, parameter discrimination, and speech intelligibility in the presence of interferers. • **Performance with Complex Stimuli:** Measuring psychophysical performance of normal-hearing and hearing-impaired listeners in several types of tasks in complex acoustical environments. • **Signal Processing Models:** Calculation of the performance on psychophysical tasks, based on optimal use of information contained at different levels of the auditory pathway. • **Studies of Hearing Impairment:** Understanding the difficulties of hearing-impaired listeners is the primary motivation of this research. • **Virtual Acoustic Environments:** Development of capabilities for spatializing sound properly in simple and complex environments. Simulation of both anechoic and reverberant environments that include single or multiple sources. These sources can be static in space or spatially-dynamic. The simulations allow exploration of what information is perceptually important for the psychophysical performance under these conditions. Exploration of components of spatialization, including simple models of basic binaural cues, empirical Head Related Transfer Functions (HRTFs), spatial interpolation of HRTFs, simple reflections, and room reverberation.	
Centre for the Neural Basis of Hearing (CNBH)	The CNBH is a new joint venture of the Physiology Department of Cam-	• **Pinna Models.** • **Physical Model of the Human Concha.**	**Ray Meddis, Lowel O'Mard** and **Chris Plack** work in this research group

University of Essex (UK) http://www.essex.ac.uk/psychology/speechlab/ *July 2006*	bridge University, the U.K. Medical Research Council, and the Psychology Department of Essex University, in collaboration with the Wellcome Department of Cognitive Neurology, University College, London. They are specialized in auditory modelling using **DSAM (Development System for Auditory Modelling**, hybrid of Matlab and AMS)	• **Absolute judgement of the location of auditory sound sources using mannequin recordings:** Using recordings made with a KEMAR mannequin, they made four different experiments to assess the accuracy of localization of auditory sound sources. • **Sound Localization in VR Systems.**	(**Trevor Shackleton** has been working here). BIBLIOGRAPHY: *Shackleton (1994).*
CIPIC Interface Laboratory UCDavis University of California http://interface.cipic.ucdavis.edu/ *July 2007*	The CIPIC Interface Laboratory is a multi-disciplinary research laboratory that brings together experts from different areas of engineering, signal processing and psychophysics. The research activities at the laboratory are concerned with human perception and its role in the interface between humans and machines. The CIPIC work focuses on spatial hearing and three-dimensional sound synthesis, with related activities in image perception, and in speech, audio, and image signal processing.	• **HRTF synthesis and customization**: Base HRTF models on underlying physics; identify the perceptually important properties; base HRTF customization on objective procedures. • **HRTF models and model composition:** Exploit model composition to structure research, identify critical parameters, and make numerical methods feasible; understand pinna/head/torso/room interactions. • **Spatial sound capture:** Capture sounds as perceived by humans; account for critical dynamic effects of motion: head rotation, body translation. • **Understand relative importance of cues:** ITD, ILD, pinna effects, torso effects, motion, room acoustics. • **Develop cost-effective HRTF customization procedures:** Parameters extracted from anthropometry; parameters extracted from imagery. • **Increase auditory realism/discrimination**: Live spatial sound capture and reproduction; virtual auditory space; augmented reality.	The CIPIC research is really important and innovative inmany respects: • Physically based model of HRTFs extracting a small number of parameters about the shape of the pinna, the shape of the head and torso (to extract these parameters, they have measured the response of an isolated pinna, see *Duda*, *2001* and *Algazi, 2002*). • Studies into the importance of the localization cues, also one separate frmo the other. • Cost-effective HRTF customization. • HRTF database with more than 90 subjects plus a Kemar mannequin (they used the Golay-code Signal with a modified Snapshot system from Crystal River Engineering), and about 2500 measurements for each subject (no distance measurements).

		• **Expand the spatial/temporal resolution for spatial sound output:** Exploit continuous models, ability to control azimuth, elevation and range. • **Demonstrate value of spatial sound for auditory interfaces:** Increase effectiveness of mobile systems, training systems, teleconferencing, monitoring and site security; more expressive musical technology. • **MTB, Motion Tracker Binaural sound**: System for the recording, storing and reproduction of binaural sound, with head motion tracked functions (the soundscape can change following the movement of the head); see *Algazi, 2005.*	BIBLIOGRAPHY: *Algazi (2001a, 2001b, 2002 and 2005), Duda (2001), Hom (2006)*
Csound Project by Eli Breder and David McIntyre http://kevindumpscore.com/docs/csound-manual/hrtfer.html *Mar 2007*	Hrtfer unit for Csound, for the dynamic binaural spatialization.	These unit generators place a mono input signal in a virtual 3D space around the listener by convolving the input with the appropriate HRTF data specified by the opcode's azimuth and elevation values. Hrtfer allows these values to be k-values, allowing for dynamic spatialization.	There are a few particular things about this object: it is said that the hrtfer makes a convolution with an HRTF database, but the object itself is just a few Kbytes, and it is improbable that it also contains an HRTF database…
CSULA Psychoacoutics Web Page California State University, Los Angeles http://www.calstatela.edu/faculty/dperrot/index.html *July 2006*	The CSULA Psychoacoustic is a laboratory inside the Department of Psychology. The main research topic of this laboratory seems to focus on the relationships between the auditory and visual systems for sound source localization. There is also a research field iton the study of localization cues (ITD, IID and DDF).	• **Visual Search** • **Loudness Matching** • **Interaural Axis** • **Apparent Motion** • **IPD, ILD Comparison** • **IPD, ITD Comparison** • **Helmet Research** • **Visual/Auditory Dominance** • **Auditory Motion Discrimination** • **Concurrent Sound Localization** • **Precedence Effect** • **Observer Weighting of Echo and Source Clicks**	These two links are really interesting because many references to the Aurally Detected Visual Search and to the Auditory Motions can be found. **http://www.calstatela.edu/faculty/dperrot/pub1.html** **David R. Perrot** is the director of this research group. BIBLIOGRAPHY: *Perrot (1997), Perrot (1993)*

Jérôme Daniel's 3D Sound Research: The Experimenter Corner (Hearing Higher Order Ambisonics) http://gyronymo.free.fr/audio3D/the_experimenter_corner.html *July 2007*	The Perceptive Aspects of Ambisonic and Stereophonic Rendering. The research focuses on the "conversion" between high order Ambisonic and binaural. Jerome Daniel now works for Orange Labs in Rennes, France, always on research topics related to Higher Order Ambisonics	• **Virtual Rendering:** This is a "light version" of the now well known "virtual loudspeaker" principle: one pair of filters (HRTF: Head Related Transfer Functions) is associated to each virtual loudspeaker as a function of its direction (from the listener's perspective), thus the signal to be delivered by the loudspeaker is filtered to yield its contribution to the left and right ear signals. The filters used are "dry" HRTF; they have been measured by K.Martin & B.Gardner on a KEMAR dummy head and have been made available for years by their authors on the MIT Web Site.	From the Internet site it is possible to download the Ph.D. thesis of Jerome Daniel, and the PowerPoint presentation. BIBLIOGRAPHY: *Daniel (2003)*
DIVA - Digital Interactive Virtual AcousticsTelecommunication Software and Multimedia Laboratory, Helsinki University of Technology. http://www.tml.tkk.fi/Research/DIVA/past/ *Jan 2009*	Diva is a collaborative research group of various topics, starting from works on the virtual reality. In their Virtual Orchestra (a virtual environment with four virtual musicians), they have implemented a 3D audio technology for headphone listening using HRTFs.	• **Real-time automatic character animation** • **Interaction through motion analysis, especially conductor following.** • **Sound generation with physical instrument models** • **Acoustics modelling and auralization** • **EVE - The Experimental Virtual Environment**	Since 2000, the DIVA personnel and projects have moved over to the new EVE Virtual Reality group (http://eve.hut.fi/), inside the HUT Lab BIBLIOGRAPHY: *Huopaniemi (1999), Savioja (1999)*
R. O. Duda Research Sound Localization Research, College of Engineering, San José State University http://www-engr.sjsu.edu/~duda/Duda.Research.html *July 2006*	This research of the San José State University is in collaboration with the CIPIC Interface Lab of the University of California at Davis. The ultimate goal of this research is to be able to build machines that can hear. The specific goal is to understand how one can locate multiple, real sound sources in natural environments.	• **Analysis of experimentally measured head-related impulse responses** to identify the primary directional cues. • **Synthesis of binaural sounds** that use these cues to create vivid, convincing images of spatial location. • **Development of computational models of the human sound localization process** that can produce accurate azimuth/elevation/range maps in	The DUDA research in really important especially for the binaural sinthesis "stage", and for the use of the localization cues by computer to establish the position of the sound source (creation of a General Sound Localization Model). BIBLIOGRAPHY: *Duda (1993), Duda (2001)*

		the presence of multiple sources and room echoes and reverberation. • **Improved human/computer interfaces:** Speech recognition systems that can distinguish talkers from each other and from environmental sounds, sound synthesis systems for spatialized auditory presentations for aircraft pilots, advanced teleconferencing systems, 3-D games, virtual reality systems, and scientific and business data display. • **Improved sound analysis systems** (such as diagnostic or monitoring systems) that can operate in reverberant, multisource environments. • **Improved binaural hearing aids** and other aids for the hearing impaired.	
ECEL Electrical and Computer Engineering Computer Lab University of Florida http://www.ecel.ufl.edu/~shassan/courses/eel6539/ *July 2006*	Matlab Implementation of 3D Synthetic Environment It's an educational project of the ECEL to show a particular algorithm implementation of Matlab. The goal is to "show how a monaural sound can be used to synthesize to come from any particular direction in 3D space".	The software, developed using Matlab, is really basic, and it seems to perform a basic convolution between the sound that needs to be spatialized and the impulse response (they are using the MIT HRTF; *see* Gardner 1994), with neither distance simulation nor HRIR interpolation techniques.	The Internet site allows listening to some spatialized samples, but they are not at all impressive. It may be noticed that no HRIR interpolation techniques are implemented, just because during the movements of the virtual sound sources clicks are clearly audible. The only strongpoint is that the interface is really simple and clear, even if not "attractive", but it seems at least to be a good implementation using Matlab.
Angelo Farina's Home Page http://pcfarina.eng.unipr.it/ *May 2009*	Main research activity: Acoustics. In more details: concert halls, musical instruments, subjective preference, auralization, numerical models for large rooms (pyramid tracing), small cavities (finite elements) and outdoors	• **Ramsete:** A powerful computer programme (based on the innovative Pyramid Tracing algorithm) for the simulation of the acoustics of closed spaces, which can also be used outdoors. • **Aurora:** A software tool for measuring, filtering and convolving the Impulse Responses of theatres	Farina's research focuses mainly on architectural acoustics. He also directs a team for the Head Related Room Impluse Response of theater: they use a system composed of two stereo microphones, a B-Format

	(image sources). Advanced measurement techniques including MLS, modal analysis, logarithmic sweep. Recently , too, DSP implementations, Underwater Acoustics and stretched-pulse measurements.	and other spaces. • **ASK Automotive Industries**: Improvement incar sound systems by means of the Auralization technique.	Soundfield microphone and a Kemar dummy head for the measurement of the impulse responses of various theatres. They use the sweep technique for the IR measurement (Aurora directly implement that technique). He also did research into the Stereo Dipole, Ambophonic and Ambisonic reproduction and synthesis techniques. BIBLIOGRAPHY: *Farina (2007 and 2009)*
HDRL: Hearing Development Research Laboratory University of Wisconsin-Madison, Fred Wightman , Doris Kistler & Co. Waisman Center http://www.waisman.wisc.edu/hdrl/index.html *July 2006*	The Hearing Development Research Laboratory (HDRL), located at the Waisman Center of the University of Wisconsin, is a small group of faculty, staff and students who conduct research into the mechanisms and processes of human hearing.	• **Development of Auditory Skills by Children**: One important result of this effort is the finding that while preschool children typically perform more poorly than adults on auditory tasks this performance is not likely to be a result of an underdeveloped auditory system, but rather a consequence of suboptimal listening strategies. • **Spatial Hearing:** Spatial hearing research has been supported by NASA as well as grants from NIH (NIDCD), the Office of Naval Research , and Rockwell Semiconductor Systems. HDRL has also received intramural support from the Center for Human Performance and Complex Systems at the University of Wisconsin - Madison. In connection with the spatial hearing research the HDRL has developed and extensively evaluated techniques for synthesizing 3-D sound images over headphones. This has stimulated both basic studies of spatial hearing processes and applied studies of potential	The HDRL Internet site allows listening to audio recorded through different dummy head (or different subjects: unstated) with different pinna shapes. The unique problem is that the audio recorded is just something like noise bursts, or short MLS or Golay Code signals, thus it is difficult properly to understand where they come from… The positive thing is that the signal is uncompressed, so the three-dimensional sensation should not be compromised by a the perceptual compression (as it happens, for example, with MP3). **Frederic Wightman, Pavel Zahoric** and **Chiron Stevens** work in this research project, and Ewan Macpherson used to.

		applications of 3-D sound technology. • **HRTF data:** They performed an experiment into the measurement of HRTFs.	The HDRL no longer exists. Wightman has now moved to the University of Louisvillle. BIBLIOGRAPHY: *Langendijk (1999; 2001), Zahorik (2001).*
The Hearing Robot by Jie Huang, University of Aizu http://www.u-aizu.ac.jp/~j-huang/Robotics/robotics.html *July 2006*	The ultimate goal of this project will be the development of a multimedia interactive autonomous mobile robot. Within the first phase development, the robot will be able to localize and track sound sources in any complex environment, have the abilities of separating target sound from environment noise, and understanding environment by sounds, e.g., human voices, sound of walking, sirens and crashes. The robot will be able to avoid obstacles and move to a destination autonomously. Visual and auditory interface will be provided.	• **Sound Localization in Reverberant Environments**: Techniques for echo cancellation for the automatic localization of the direction of the sound sources in a three-dimensional space. • **Auditory Scene Analysis:** Talking to other persons in a noisy room, walking in the street, the auditory system faces the problem of separating complex sound signals into different sound streams. • **Sound Understanding:** Sound understanding is important for an autonomous robot to recognize its environment.	The main goal of this research is the development of an automatic sound source localization (by computer), using the localization cues, techniques for echo suppression, and others. BIBLIOGRAPHY: *Huang (1997), Huang (1999).*
HRTF Measurements of a KEMAR Dummy-Head Microphone by Bill Gardner and Keith Martin, MIT Media Lab http://sound.media.mit.edu/resources/KEMAR.html	An extensive set of head-related transfer function (HRTF) measurements of a KEMAR dummy head microphone was completed in May, 1994.	• The measurements consist of: the left and right ear impulse responses from a Realistic Optimus Pro 7 loudspeaker mounted 1.4 metres from the KEMAR. • Signal used for the HRIR extraction: Maximum length (ML) pseudo-random binary sequences (MLS) were used to obtain the impulse responses	At the MIT, there is also the **Research Laboratory of Electronics**, with its**Auditory Perception and Cognition Group**, and the **Speech and Hearing Bioscience and Technology Programme**: both deal with binaural research.

July 2006		at a sampling rate of 44.1 kHz. A total of 710 different positions were sampled at elevations from -40 degrees to +90 degrees. • Other measurements: The impulse response of the speaker in free field and several headphones placed on the KEMAR.	**Bill Gardner** and **Keith Martin** work in this research group. BIBLIOGRAPHY: *Gardner (1994).*
Human Interface Technology Laboratory HITL University of Washington http://www.hitl.washington.edu/ *Jan 2009*	The Human Interface Technology Lab (HITLab) is a multi-disciplinary research and development lab whose work centers around human interface technology. Lab researchers represent a wide range of departments from across the University of Washington campus, including engineering, medicine, education, social sciences, architecture and the design arts.	In the binaural field, they are mainly developing systems for the Virtual Reality (the main part of the research is about video), with binaural spatialization function. They are also developing the VRD, a three-dimensional display technology for US Navy pilots.	BIBLIOGRAPHY: *Winn (2005).*
HUT Acoustics Lab Laboratory of Acoustic and Audio Signal Processing Helsinki University of Technology (TKK) http://www.acoustics.hut.fi/ *June 2009*	The TKK Laboratory of Acoustics and Audio Signal Processing is the only university unit in Finland that focuses primarily on research into and the teaching of acoustics, speech, and sound processing. The laboratory was established in its current form in 1981.	• **Analysis and sinthesys of musical sounds:** Most of the music-related research conducted at our laboratory is related to sound synthesis by physical modelling. • **Special digital filter for audio signal processing**. • **Speech modelling and enhancement:** Mathematical modelling of speech, its theory and applications in glottal inverse filtering and in speech transmission. • **Voice Production Analysis:** Analysis and parametrisation of the voice source in speech and singing, with application to occupational voice research and cognitive brain research. • Communication acoustics team, spatial sound and psychoacoustics: Spatial sound reproduction and	**Ville Pulkki, J. Merimaa, T. Lokki, T. Hirvonen, L. Savioja** and **Matti Karjalainen** work in this research group. BIBLIOGRAPHY: *Ahonen (2009), Pulkki (1997-2001; 2002; 2005; 2006), Merimaa (2002; 2004; 2005), Hirvonen (2004), Lokki (2002), Tikander (2004), Karjalainen (2004; 2005).*

		real and virtual acoustics. Psychoacoustics of spatial sound and musical instruments. In this research field, the research topics are: • Vector Based Amplitude Panning (VBAP): A widely used technique for positioning virtual sources using amplitude panning for arbitrary 2-D or 3-D loudspeaker layouts. • Spatial Impulse Response Rendering (SIRR): A recent technique for introducing the acoustic effect of an existing room to a dry sound signal over multi-channel loudspeaker setups. The impulse response of SIRR is commercially available from Waves Inc. • Directional Audio Coding (DirAC): A technique for the encoding and decoding of 3D sound fields, mainly used for teleconferencing. • Psychoacoustic and Evaluation of Spatial Sound: Listening tests and binaural auditory modelling. • Real and Virtual Acoustics: Analysis of sound field in spaces and room simulation techniques for interactive virtual acoustics (most all-way tracing techniques). • Augmented Reality Audio: The concept of augmented reality audio (ARA) characterizes techniques where a physically real sound and voice environment is extended with various virtual environments and	

		communication scenarios. In this project, the focus is on developing new techniques, applications, and interfaces for future wearable personal appliances which support full-duplex high-quality audio and voice transmission. • DSP techniques for room response modelling: Presents a particular digital filter configuration and related design methods for room response modelling and equalization. • Perception of musical instrument sounds: Perception of attributes in musical instrument sounds, and subjective evaluation of sound source models.	
Immersive Audio Lab Integrated Media Systems Center, University of Southern California - Chris Kyriakakis, Tomlinson Holman imsc.usc.edu/research/project/immersiveaudio/immersiveaudio_tech.pdf *July 2006*	Research in the IMSC Immersive Audio Laboratory focuses on algorithms for capturing and rendering sound so that it is indistinguishable from reality. Our goal is to provide an immersive experience through greater imaging and envelopment capabilities than ever before.	• **Virtual Microphone**: (convert mono or stereo recordings to multichannel immersive audio in real time). A technology that uses sophisticated algorithms and filters to unlock the multichannel audio potential of one- and two-channel recordings. Reported by Chris Kyriakakis: "From a real recording, even if you're close to the orchestra, you're still going to have some reverberation caused by the room's acoustics. Adding reverb on top of original reverb is what causes audible artifacts. We have the visual equivalent of morphing images: Take one signal and transform it into the other. In some cases, we add reverb in certain frequencies, and in others we subtract. It's not just run through a reverberator to get a surround signal." • **10.2 Immersive Audio Rendering:** Multichannel	It was impossible to find an Internet site of the laboratory: all of the information about this isextracted from reviews and articles and from a PDF document with a description of the lab (*see* link in the first window). **Tomlinson Holman** and **Chris Kyriakakis** work in this research group. BIBLIOGRAPHY: *Bharitkar (2003), Mouchtaris (2003), Kyriakakis (2000).*

		rendering algorithms that use acoustics, psycho-acoustics, and adaptive audio signal processing to immerse a group of listeners in a seamless sonic environment. • **Dynamic spatialization**: Software for moving sound seamlessly using five, ten, or more loud-speakers, compatible with industry-standard digital audio workstation software.	
Institute of Electronic Music and Acoustic (IEM) Graz, Austria http://iem.at//index_html_en *Jan 2009*	Various work in room acoustics simulation using various techniques (Wavefield, Ambisonic…), and they are a really active centre for the development of Pure Data objects. They also work on other acoustics topics, such as beamforming applica-tions…	They developed various Pure Data objects for room acoustics simulation. An example of these imple-menting the binaural technique could be the bin_ambi.OSC, a Sound Server Application for Real-Time Binaural Audio Rendering in Pure Data (PD) (2006-present). They are also involved in different research projects related to High Order Ambisonic.	**Markus Noisternig** (now at IRCAM)**, A. Sontacchi, F. Zotter and T. Musil** work in this research group. BIBLIOGRAPHY: *Zotter (2007), Musil (2005), Noisternig (2003).*

Institute of Communication Acous-tics
Ruhr Universitat Bochum
http://www.ika.ruhr-uni-bochum.de/ika/forschung/gruppe_blauert/asas_eng.htm#binaural
 	
Jan 2009
	Realizing a comfortable acoustical man-machine interface, where the machine can be either the source or the sink of information, is highly desirable. While binaural technology focuses on both of the directions in information processing, the main objective of the binaural-technology research group is to develop systems that process acoustically transmitted information (i.e. the sink of information). This is done by using humans as models for the technical systems, such as hearing aids, speech recognizers, hands-free communication systems, teleconferen-	• **Analysis and Synthesis of Auditory Scenes (ASAS):**
		  • Auditory Signal Processing & Binaural Technology (Jens Blauert, Juha Merimaa, Wolfgang Hess): research into machines such as the "Cocktail Party Processor", capable of extracting the desired acoustic information (i.e., a speaker's utterance) from disturbing background noise.
		  • Simulation and Virtual Environments (M. Ercan Altinsoy, Jens Blauert, Antonio Guz-man Avalos, Pedro Novo, Andreas Silze): The main research activities deal with prob-lems related to simulating listening in reflective environments, since sound reflected
			Really important research centre on the physiological and psychoacoustic aspects of spatial audio perception.
			Jens Blauert (*see* Blauert, 1996) works in this research group.

	cing systems, localization systems and binaural measurement and analysis systems (measurements with dummy heads), to be designed. To achieve this, the human auditory system is investigated psychoacustically and neurophysiologically.	off walls has a dominant influence on the auditory spatial impression.	
IRCAM Room Acoustic Team Directed by Olivier Warusfel http://www.ircam.fr/salles.html?&L=1 *Jan 2009*	The research and development activity carried out by the Room Acoustics Team concentrates on the analysis, reproduction and synthesis of spatially considered sound environments. In other words, the localization of sound in space and the specific acoustic characteristics of a room or environment. The goal is to provide models and tools, which will enable composers to integrate the spatial organization of sound into their work from its conception through to the concert situation.	• **Auditory spatial cognition**: In the field of sound reproduction and communication, future audio technology will attempt to shift emphasis towards sensations of immersion and presence. Such notions are intimately linked to the spatial dimensions of a multimedia scene and are intensified in situations involving the participation of the listener. This participation may involve navigation within a scene or the gestural interaction of objects within it. • **Authoring tools for spatialization:** The processes involved in spatialization demand the development of authoring tools to edit sound scenes and manipulate them in real time. These editing tools are based on models of the physical or perceptive description of a sound scene. The data supplied are transferred to the spatialization motor *via* a communication protocol. As part of the CARROUSO and LISTEN European Projects, **ListenSpace** software was adapted to work with an encoding-transmission chain, formatted in MPEG-4. • **LISTEN European Project:** Extraction of an HRTF database from 51 subjects. • **Automatic Extraction of Spatial Descriptors:**	Inside the Forum IRCAM software packaging, they are developing a asoftware called SPAT for the multichannel and binaural spatialization. Inside SPAT they have implemented a binaural technique, a HRIR interpolation technique and a reverb technique. The HRTF are extracted from subjects (no dummy head) in an anechoic chamber. **Olivier Warusfel, Terence Caulkins, Markus Noisternig, Etienne Corteel** (now Sonic Emotions)**, Brian Katz** (now LIMSI-CNRS) and **Suzanne Winsberg** work in this research group. BIBLIOGRAPHY: *Nguyen (2009), Corteel (2007), Tsingos (2006), Warusfel (2004), Jot (1999).*

		Initially, work started out by developing automatic methods for the objective description of a recorded sound scene's spatial qualities, with no prior knowledge of the sources it contains, the sound message given off or the nature of the recording space. Study now focuses on an attempting to estimate the source's direction and the characteristics of the reverb envelope. A theoretical framework, based on spatial audition models, allows a homogenous set of detection/estimation methods to be developed. The information obtained in each frequency band is then collected and interpreted using higher-level descriptor. • **Binaural Reproduction Technology:** This technique is based on a dynamic filtering of the sound source using transfer functions (HRTF - Head Related Transfer Functions) measured on the head of a listener or model. • **Hopophonic Sound Field Synthesys (european CARROUSO research project):** Wave Field Synthesis (WFS) is a holophonic reproduction process, which enables a sound scene to be captured or synthesized, by analogy to visual holograms, whilst preserving spatial characteristics of distance and source direction. The accuracy of reproduction of the extended listening zone made possible by this approach, initiated by Delft University, goes well beyond the limitations of conventional systems. Whereas conventional stereophonic techniques (stereo 5.1) can only be appreciated from the centre of the listening zone,	

		the aim of holophonic technology is to reproduce a sound field in which listeners can maintain a coherent perception of the perspectives between sources whilst moving around. • **Wavefield synthesis** • **SPAT:** MaxMSP objects library for sound spatialization with binaural functions.	
Gary S. Kendall Assoc. Professor of Music Technology and Music Theory - Northwestern University http://musictechnology.northwestern.edu/3dsound/ *Jan 2009*	Northwestern's 3D Sound and binaural recording pages, courtesy of Professor Gary Kendall and students.	• **Binaural Northwestern**: Developed by Adam Fenton, it is a collection of binaural recordings from locations on the Northwestern University Evanston Campus that incorporate QTVR movies of each location. • **Binaural Chicago**: Take a sound guided tour of Chicago hotspots with the website Binaural Chicago, created by Cari Morin, a graduate of the Masters in Music Technology programme. • **HRTF Display:** HRTF Visual Display is a Flash application for viewing graphic data sets that relate to Head Related Transfer Functions. Developed as a Master of Music project by Adam Fenton, HRTF Visual Display allows a user to choose Impulse Response locations and then automatically displays the corresponding visual data. • **3D Sound Primer:** A tutorial developed by Kendal about 3D audio (**http://musictechnology.northwestern.edu/3dsound/primertop.html**).	It is hard to find information on HRTF extractions, such as the signal used, the technique, or the azimuth and elevation scale. The HRTF display is really useful, most of all for the ITD diagrams (group delays). The 3D Sound Primer is a particularly simple and clear tutorial 3D audio. BIBLIOGRAPHY: *Kendall (1982; 1984; 1988; 1992).*
LabROSA Columbia University, Department of Electrical Engineering,	The Laboratory for the Recognition and Organization of Speech and Audio (LabROSA) conducts research into	There are many research interests in this Lab, and those related to spatial audio are: • Computational auditory scene analysis to	The research is mainly based on the automatic recognition of speech (including spatial aspects linked to it).

http://labrosa.ee.columbia.edu/ *July 2006*	automatic means of extracting useful information from sound. Our vision is of an intelligent 'machine listener', able to interpret live or recorded sound in terms of the descriptions and abstractions that would make sense to a human listener.	decompose sound mixtures according to the attributes of the individual sources, and the contribution of the acoustic environment • Speaker identification and characterization.	
LIMSI-CNRS University of Paris Sud (Paris 11) Audio Acoustique Group, Sound & Space http://www.limsi.fr/Scientifique/ps/thmsonesp/SonEspace *Jan 2009*	Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur. The team Sound & Space conducts various research into room acoustic simulation, 3D sounds through loudspeakers using Ambisonic, wavefield sinthesys and binaural technques. They frequently collaborate with IRCAM and with the group of Olivier Warusfel.	They developed various engines for the binaural spatialization of soundfiles (for example, the LSE, Limsi Spatialization Engine, with a technique for ITD customization and ILD modelling for close sound sources) and for the conversion between Ambisonic and binaural using virtual loudspeakers. They work on research projects involving: • Soundscape perception by blind people • Directional sound sources (simulation of the directionality of a talking head) • HRTF individualization and selection. A joint research project between LIMSI and Arkamys (www.arkamys.com) has been activated in 2007, with the topic of binaural systems with HRTF customization for mobile devices.	**Brian Katz, Markus Noisternig** (now IRCAM)**, Emmanuel Rio, David Shonstein, Nick Mariette and Lorenzo Picinali** (now De Montfort University, Leicester) work in this research group. BIBLIOGRAPHY: *Katz (2004), Afonso (2005).*
Ewan Macpherson's Research Central Systems Laboratory at the Kresge Hearing Research Institute, University of Michigan http://www-personal.umich.edu/~emacpher/ *July 2006*	Directly from the internet site: *My main interest is in the mechanisms of sound localization, and in particular the processing of "pinna" or spectral cues. The aim is to understand how the auditory system generates a percept of a sound source situated in space from the signals reaching the eardrums. Or, as my mother puts it when explaining*	• **Computer models for binaural localization**	The Macpherson work is primarily focused on the physiological aspects of the directional hearing, then on the computer simulation of these. BIBLIOGRAPHY: *Macpherson (1991; 2002; 2003), Middlebrooks (2000), Sabin (2005).*

	what I do: How do we hear in three dimensions when we have only two ears?		
William Martens spatial hearing research, McGill University, Schulich School of Music, http://www.music.mcgill.ca/~wlm/ *July 2006*	He is a perceptual psychologist specializing in spatial hearing research and the simulation of the acoustical cues used in human sound localization. He holds several patents on spatial sound processing technology and has contributed to the development of several commercial spatial sound processing technologies.	• **HRTFs simulations and generalized HRTFs** • **Perceptual evaluations of HRTF filtering**	BIBLIOGRAPHY: *Martens (2003)*
Keith Martin's Research Interests MIT Media Lab http://alumni.media.mit.edu/~kdm/research.html *July 2006*	Research interest of Keith Martin in the Spatial Hearing Field	• Cones of cunfusion • **Head-related transfer functions.** • **Spatial head model**: By modelling the precedence effect, and carefully decoding the interaural time and intensity differences between the signals at the two ears, it should be possible to construct a spatial hearing model that performs similarly to human listeners.	
NRC, Nokia Research Center Various labs around the world (Helsinki, Palo Alto, Budapest…) http://research.nokia.com/research/index.html *July 2006*	Prototype implementation of an algorithm for the binaural and transaural spatialization	• It is not possible to gain much information from the NRC internet site; all of the information I have about this research group (obviously just in the binaural spatialization field) is in *Lorho (2004)*.	BIBLIOGRAPHY: *Lorho (2004)*
Parmly Hearing Institute Loyola University Chicago http://www.parmly.luc.edu/	Parmly is a basic science research institute of the Graduate School of Loyola University Chicago engaged in the comparative study of sensory	• **Behavioral and Psychophysical Research:** The major objective of behavioural and psychophysical research is to characterize the ability of various sensory systems to perform complex signal analy-	The research areas of this group mainly focus on the physiological and psychological aspects of the hearing system (especially on the inner part of the

Jan 2009	systems including hearing, vision, speech perception, vestibular function, and the special senses of the lateral-line organ and electroreception in fish and humans.	sis in order for animals to determine stimulus sources in their environment. • **Physiological and Anatomical Research.** • **Modelling in Research and Education:** Modelling, simulations, and data bases are used and developed at Parmly in several areas: • **Simulations (Auditory Virtual Space):** Auditory virtual space means the ability to experience a real world auditory environment through simulations presented over loud-speakers or headphones. For instance, being able to experience listening to a concert at Orchestra Hall while sitting in your living room, or an air show while lying in bed (the simulation is done through analyzing the re-sponse of a Kemar mannequin). • **Neural Modelling (Auditory Image Model):** The inner ear and the auditory periphery take in sound and convert it to a set of neural sig-nals that are relayed to the brainstem and brain. These neural signals form a neural code of the original sound. The neural circuits of the brainstem and brain further process this code to determine what and where the objects that produced the sound are located.	hearing system, on the auditory periph-ery, and on the brainstem). **Richard R. Fay, William A. Yost, Stan Sheft, R. H. Dye, David Green** and **Terry Grande** works in this research group. BIBLIOGRAPHY: *Yost (1970; 1972; 1975; 1997), Green (1975), Dye (1996), Sheft (1997).*
Ville Pulkki's Research Interests Laboratory of Acoustic and Audio Signal Processing Helsinky University of Technology	Research interest of Ville Pulkki in the Spatial Hearing Field (*see* HUT)	• **Multi-channel 3-D sound using Vector Base Amplitude Panning VBAP (*see* HUT)** • **Evaluation of spatial sound quality** • **Virtual acoustics and analysis of real room**	From the personal page of Ville Pulkki can be downloaded the VBAP software for Csound, MAX-MSP, Pure Data, etc. It is also possible also to download some

http://www.acoustics.hut.fi/~ville/ *Jan 2009*		**acoustics** • **Visualization of edge diffraction** • **Spatial Impulse Response Rendering** • **DirAC (see HUT)**	demos of the Spatial Impulse Response Rendering.
Sheffield Speech and Hearing Research Group (SPANDH) The University of Sheffield, Department of Computer Science http://www.dcs.shef.ac.uk/spandh/ *Jan 2009*	The Speech and Hearing Research Group (SpandH) was established in the Department of Computer Science, University of Sheffield, in 1986. The group is concerned with computational modelling of auditory and speech perception in humans and machines, robustness in speech recognition and large vocabulary speech recognition systems and their applications.	• **Computational Auditory Scene Analysis.** • **Missing Data methods for robust automatic speech recognition.** • **Large Vocabulary Continuous Speech Recognition.** • **Information Retrieval from Spoken Corpora.** • **CAST - Clinical Applications of Speech Technology.** The fouded research project on the binaural area are: • **Incorporating Binaural Cues in a Computational Model of Auditory Scene Analysis**: Auditory Scene Analysis (ASA) promises to provide the needed front-end for robust automatic speech recognition devices. A more powerful approach to the current monaural one would be to include binaural cues. Listeners are able to use such cues as timing and intensity differences between the two ears to locate sounds in space, and to group sounds that originate from the same spatial location. This project aims to model this process in a physiologically plausible manner. • **Auditory Display**: Ph.D. research project.	This is not properly a research centre in the binaural field: they are working on research fields linked to hearing and speech, with some binaural functions for the Auditory Display and for the Auditory Scene Analysis.
Spatial Audio Work in the GVU MCG Georgia Institute of Technology http://www-static.cc.gatech.edu/gvu/multimedia/sp	A research group into the spatial audio field, inside the Multimedia Computing Group. They work with HRTF, although did not extract their own database: they use	• Environmental Modelling for binaural audio • Software engineering for binaural audio engines	BIBLIOGRAPHY: *Burgess (1992)*

atsound/spatsound.html *July 2006*	the MIT database.		
Spatial Media Group at the University of Aizu by Michael Cohen - Pioneer Sound Field Control System http://www.u-aizu.ac.jp/~mcohen/spatial-media/welcome/2005.html *July 2006*	Various tasks in the Spatial Audio Field, especially for the commercialization of products for home entertainment, for mobile phones and virtual reality.		On the group's Internet site, I could find little information on binaural studies.
The Virtual Acoustics Project in the ISVRThe Institute of Sound and Vibration Research (ISVR) at the University of Southampton http://www.isvr.soton.ac.uk/FDAG/vap/ *Jan 2009*	Improve the ability of audio systems to produce "images" of sound sources perceived by the listener. *"We try to produce the illusion in a listener of being in a 'virtual' acoustic environment which is entirely different from that of the space in which he (or she) is actually located".*	• **Virtual Source Imaging** • **Stereo Dipole and Cross Talk Cancellation Techniques** • **Multi Channel signal processing techniques** • **Visually Adaptive Imaging**: selection of appropriate virtual audio filters that correspond to a listener's varying head position. • **3D models of pinna and dummy head** • **BEM simulation of ellipsoid HRTFs (Matlab workspace):** This is an HRTF database calculated from an ellipsoid simulation of the head at 19 elevation angles, at 72 azimuth angles, with 51 frequencies and at three distances (25cm, 1m, 10m). • **HRTF measurements:** using MLSSA (Maximum Length Sequence System Analyzer) and MLS signal technique. • **High Order Ambisonic**	• This research is joined to that of Professor Hareo Hamada's group at Tokyo Denki University (TDU) in Japan. • They are also working on the influence of individual head related transfer functions. • The next step of the HRTF research is to obtain high accuracy IR using other methods for extending the frequency range. • On the Internet site there are many useful animations for, e.g., cross-talk cancellation, and the free field model with a Kemar dummy head. There are also some downloads for the Stereo Dipole wave fiiles, HRTF database, HRTF simulation. **P. Nelson, F. Fazi** and **O. Kirkeby**

			work in this lab. BIBLIOGRAPHY (the publication list on their internet site is updated to 2001): *Nelson (1995; 1996), Kahana (1997), Takeuchi (1997a; 1997b).*

Appendix Bibliography

[1] Abel, J. S. and Foster, S. H., *Snapshot HRTF Measurement System User's Guide*, Crystal River Engineering, 490 California Ave., Suite 200, Palo Alto, CA 94306, USA, 1994

[2] Abel, J. S. and Foster, S. H., *Measuring HRTFs in a Reflective Environment*, In G. Kramer & S. Smith (Eds.), Proceedings of the 1994 International Conference on Auditory Displays, (p. 265). Santa Fe, NM, 1995

[3] Afonso A., Katz B.F.G., Blum A., Denis, *Spatial knowledge without vision in an auditory VR environment*. M..ESCOP 2005. XIVth Conference of the European Society for Cognitive Psychology. Leiden, The Netherlands, August 31-September 3, 2005

[4] Ahonen, J., Pulkki, V., Kuech, F., Galdo, G. D., Kallinger, M., and Schultz-Amling, R., "Directional Audio Coding with stereo microphone input," in Proceedings of The 126th AES Convention, preprint 7708, pp. 1-9, Munich, Germany, May 7-10, 2009.

[5] V. R. Algazi and R. O. Duda and D. M. Thompson and C. Avedano, *The CIPIC HRTF Database*, IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2001a

[6] V. R. Algazi, C. Avendano and R. O. Duda, *Estimation of a spherical-head model from anthropometry*, J. Aud. Eng. Soc., Vol. 49, No.6, p.472-478, Junee, 2001b

[7] V. R. Algazi, R. O. Duda, R. Duraiswami, N. A. Gumerov and Z. Tang, *Approximating the head-related transfer function using simple geometric models of the head and torso*, J. Acoust. Soc. Am., Vol.112, pp. 2053-2064, Nov. 2002

[8] V. R. Algazi, R. J. Dalton, R. O. Duda and D. M. Thompson, *Motional-Tracked Binaural Sound for Personal Music Players*, Paper 6557, 119th Convention of the Audio Engineering Society, New York, NY, Oct. 2005

[9] D. R. Begault, M. R. Anderson, B. U. McClain, & J. D. Miller (2005) Audio-Visual Communication Monitoring System for Enhanced Situational Awareness, "Working Together: R&D Partnerships in Homeland Security Conference", Boston, MA, 27-28 April 2005

[10] S. Bharitkar and C. Kyriakakis, *Selective Signal Cancellation for Multiple-Listener Audio Applications Using Eigenfilters*, IEEE Transactions on Multimedia, Junee 2003

[11] Blauert, J. (1996). *Spatial Hearing, the Psychophysic of Human Sound Localization*. Cambridge, Massachussetts, USA: The MIT Press Cambridge.

[12] Burgess, D.A., *Techniques for low-cost spatial audio*, ACM Fifth Annual Symposium on User Interface Software and Technology (UIST '92), Monterey, November 1992

[13] Cockayne William and Zyda Michael and Barham Paul and Brutzman Don and Falby John, *The laboratory for human interaction in the virtual environment*, Presented at ACM Symposium on Virtual Reality Software and Technology (VRST), Honk Kong: University of Hong Kong, 1996

[14] E. Corteel « Synthesis of directional sources using Wave Field Synthesis, possibilities and limitations », EURASIP Journal on Advances in Signal Processing, special issue on Spatial Sound and Virtual Acoustics, Janvier, 2007

[15] Daniel, J., Rozenn, N., & Moreau, S. (2003). Further Investigations on High Order Ambisonic and Wavefield Synthesis for Holophonic Sound Imaging. In *Proceedings of the 114th Conference of the Audio Engineering Society, 22-25 March, Amsterdam, The Netherlands.*

[16] Richard O. Duda, *Estimating Azimuth and Elevation From the Interaural Intensity Difference*, Technical Report No. 4, NSF Grant No. IRI-9214233, Dept. of Elec. Engr., San Jose State Univ. 1993

[17] Richard O. Duda, *Anthropometry in the CIPIC HRTF Database*, CIPIC Interface Laboratory, UCDavis University of California, 2001

[18] Dye, R. H., *"The Relative Contributions of Targets and Distractors in Judgments of Laterality Based on Interaural Differences of Level," in Binaural and Spatial Hearing in Real and Virtual Environments* (eds. R. H. Gilkey and T. R. Anderson), Lawrence Erlbaum, Associates, 1996

[19] Farina, A. (2009). Silence Sweep: a novel method for measuring electro-acoustical devices. In *Proceedings of the 126 AES Convention*, Munich, Germany.

[20] Farina, A. & Farina, A. (2007). Real-Time Auralization Employing a Not-Linear, Not-Time-Invariant Convolver. In *Proceedings of the 123th AES Convention* , New York, US.

[21] Bill Gardner and Keith Martin, *HRTF Measurements of a KEMAR Dummy-Head Microphone,* MIT Media Lab Perceptual Computing - Technical Report #280, Boston, Massachussetts, USA, 1994

[22] Gilkey, R.H. and Anderson, *Binaural and Spatial Hearing in Real and Virtual Environments*, , T.R. (eds.), New Jersey: Lawrence Erlbaum Associates, 1997

[23] Green David M. and Yost William A., *Binaural Analysis, A Chapter for volume 5(2) of Handbook of Sensory Physiology: Hearing*, (Edited by Keidel and Neff), Springer-Verlag, 1975

[24] Hafter, E. R. and Trahiotis, C. (1997) "Functions of the Binaural System," in Handbook of Acoustics, M. Crocker (ed). Wiley.

[25] Hirst, J.M. and Davies, W.J. and Phiipson, P.J., *Multichannel spatialization techniques for musical synthesis*, Proc. Institute of Acoustics 22(6), 117-128, 2000

[26] T. Hirvonen and M. Tikander and V. Pulkki, *Multichannel reproduction of low frequencies*, Baltic-Nordic Acoustics Meeting, 2004 Junee 8-10 Mariehamn, Åland, Finland, 2004

[27] Pablo Faundez Hoffmann and Henrik Møller, *Audibility of spectral differences in head-related transfer functions*, To be presented at the 120th AES Convention, May 20-23, Paris, France, 2006

[28] Pablo F. Hoffmann, Henrik Møller (2008a): "Audibility of direct switching between head-related transfer functions", Acta Acustica united with Acustica, Vol. 94 (2008), pp. 955-964.

[29] Pablo F. Hoffmann, Henrik Møller (2008b): "Some observations on sensitivity to HRTF magnitude", Journal of the Audio Engineering Society, Vol. 56, No. 11, November 2008, pp. 972- 982.

[30] R. C-M Hom, V. R. Algazi and R. O. Duda, "High-frequency interpolation for motion-tracked binaural sound," paper 6963, 121st Convention of the Audio Engineering Society, San Francisco, CA (Oct. 2006)

[31] Huopaniemi J. , *Virtual Acoustics and 3-D Sound in Multimedia Signal Processing*, Doctoral Thesis, Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing, Report 53, 1999

[32] Huang J. and Ohnishi N. and Sugie N., *Building Ears for Robots: Sound Localization and Separation*, Artificial Life and Robotics (Springer-Verlag), Vol.1, No.4, pp.157-163, 1997

[33] Huang J. and Supaongprapa T. and Terakura I and Wang F. and Ohnishi N. and Sugie N., *A Model Based Sound Localization System and Its Application to Robot Navigation*, Robotics and Autonomous Systems (Elsevier Science), Vol.27, No.4, pp.199-209, 1999

[34] J. Jot, *Real-time spatial processing of sounds for music, multimedia and interactive human-computer interfaces*, Multimedia Systems, vol. 7, n° 1, Janvier, 1999

[35] Y. Kahana, *Multi-channel sound reproduction with a four-ear dummy-head*, M.Sc. thesis. Institute of Sound and Vibration Research, University of Southampton, England, 1997

[36] Karjalainen M. and Paatero T. and Mourjopoulos J. N. and Hatziantoniou P. D., *'About Room Response Equalization and Dereverberation*, in Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA'05), pp. 183-186, New Paltz, NY, USA, October 16-19, 2005

[37] M. Karjalainen, M. Tikander, and A. Härmä, *Head-tracking and subject positioning using binaural headset microphones and common modulation anchor sources*, in Proceedings of the ICASSP'2004, Montreal, Canada, 17-21 May, 2004

[38] Katz B.F.G., *International round robin on room acoustical impulse response analysis software 2004,* Acoustics Research Letters Online, Vol 5, n°4, 158-164, October, 2004

[39] Kendall, G. S. and C. A. P. Rodgers., *The simulation of three-dimensional headphones cues for headphone listening*, Proceedings of the 1982 International Computer Music Conference, 1982

[40] Kendall, G. S. and W. L. Martens, *Simulating the Cues of Spatial Hearing in Natural Environments*, Proceedings of the 1984 International Computer Music Conference, 1984

[41] Kendall, G. S., W. L. Martens, and M. D. Wilde, *A Spatial Sound Processor For Loudspeaker and Headphone Reproduction*, The Sound of Audio. Proceedings of the AES 8th International Conference, 1988

[42] Kendall, G. S., *Directional Sound Processing In Stereo Reproduction*, Proceedings of the 1992 International Computer Music Conference, pp. 261-264, 1992

[43] C. Kyriakakis, *Virtual Loudspeakers and Virtual Microphones for Multichannel Audio*, IEEE International Conference on Consumer Electronics, Los Angeles, Junee 15th, 2000

[44] Langendijk E. H. A. and Wightman F. L. and Kistler D. J., *Sound localization in the presence of one or two distracters*, Abstracts of the 22nd Midwinter Meeting, Association for Research in Otolaryngology, 1999

[45] Langendijk E.H.A. and Wightman F.L. and Kistler D.J., *Sound localization in the presence of one or two distracters*, Journal of the Acoustical Society of America, 109, 2123-2134, 2001

[46] T. Lokki and V. Pulkki*, Evaluation of geometry-based parametric auralization*, Audio Engineering Society 22th Int. Conf on Virtual, Synthetic and Entertainment Audio pp. 367-376. Junee 15-17, Espoo, Finland, 2002

[47] Gaetan Lorho and Nick Zacharov, *Subjective Evaluation of Virtual Home Theatre Sound Systems for Loudspeakers and Headphones*, Preprint of the Audio Engineering Society for the 116th Convention, Berlin, Germany, 2004

[48] Macedonia Michael R. and Brutzman Donald P. and Zyda Michael J. and Pratt David R. and Barham Paul T. and Falby John and Locke John, *NPSNET: A multi-player 3-D virtual environment over the internet*; In Proceedings of the 1995 Symposium on Interactive 3-D Graphics . Monterey, California, 1995

[49] Macpherson, E.A. *A computer model of binaural localization for stereo imaging measurement*, J. Audio. Eng. Soc., 39(9): 604-622, 1991

[50] Macpherson, E.A. & Middlebrooks, J.C., *Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited*, J. Acoust. Soc. Am., 111(5):2219 2236, 2002

[51] Macpherson, E.A. & Middlebrooks, J.C., *Vertical-plane sound localization probed with ripple-spectrum noise*, J. Acoust. Soc. Am., 114:430-445, 2003

[52] Martens, W. L., *Perceptual Evaluation of Filters Controlling Source Direction: Customized and Generalized HRTFs for Binaural Synthesis*, Acoustical Science and Technology, 24 (5), 220-232, 2003

[53] J. Merimaa and V. Pulkki, *Spatial Impulse Response Rendering I: Analysis and Synthesis*, Journal of the Audio Engineering Society, vol 53, no. 12, 2005

[54] J. Merimaa and W. Hess, *Training of Listeners for Evaluation of Spatial Attributes of Sound*, in Proceedings of The 117th AES Convention, preprint 6237, San Francisco, CA, USA, October 28-31, 2004

[55] J. Merimaa, *Applications of a 3-D Microphone Array*, presented at the 112th AES Convention, preprint 5501, Munich, Germany, May 10-13, 2002

[56] Middlebrooks, J.C., Macpherson, E.A. & Onsan, Z.A., *Psychophysical customization of directional transfer functions for virtual sound localization*, J. Acoust. Soc. Am. 108(6), 3088-3091, 2000

[57] Pauli Minnaar and Jan Plogsties and Flemming Christensen, *Directional resolution of head-related transfer functions required in binaural synthesis*, Journal of the Audio Engineering Society, Vol. 53, No. 10, October, pp. 919- 929, 2005

[58] A. Mouchtaris and S. S. Narayanan and C. Kyriakakis, *Virtual Microphones for Multichannel Audio Resynthesis*, EURASIP Journal on Applied Signal Processing, Special Issue on Digital Audio for Multimedia Communications, May 2003

[59] T. Musil, M. Noisternig, and R. Hoeldrich*, A Library for Realtime 3D Binaural Sound Reproduction in Pure Data (PD),* Proc. Int. Conf. on Digital Audio Effects (DAFX-05), Madrid, Spain, September 20-22,2005

[60] P.A. Nelson, F. Orduna-Bustamante and H. Hamada, *Inverse filter design and equalisation zones in multi-channel sound reproduction*, IEEE Transactions on Speech and Audio Processing 3 (3), pp. 185-192, 1995

[61] P.A. Nelson and F. Orduna-Bustamante, D. Engler, and H. Hamada, *Experiments on a system for the synthesis of virtual acoustic sources*, J. Audio Eng. Soc. 44, pp. 990-1007, 1996

[62] K. Nguyen, C. Suied, I. Viaud-Delmon, O. Warusfel « patial audition in a static virtual environment : the role of auditory-visual interaction », Journal of Virtual Reality and Broadcasting, 2009

[63] Noisternig, M., Musil, T., Sontacchi, A. & Hoeldrich, R. (2003). *3D Binaural Sound Reproduction using a Virtual Ambisonics Approach*. Proc. Int. Symp. on Virt. Env., Human-Computer Interf., and Meas. Sys. (VECIMS), Lugano, Switzerland, Julyy.

[64] Perrott D.R., *Auditory and visual localization: Two modalities and one world*, Proceedingsof the 12th International Conference of the Audio Engineering Society, Snekkersten, Copenhagen, Denmark, 1993

[65] Perrott, D.R. and Strybel, T.Z., *Some observation regarding motion-without-direction*, In Binaural & Spatial Hearing, edited by T. Anderson and R. Gilkey, Lawrence Erlbaum Associates, Publishers, Mahwah, New Jersey, pgs.275-294, 1997

[66] V. Pulkki and J. Merimaa, *Spatial Impulse Response Rendering II: Reproduction of Diffuse Sound and Listening Tests*, Journal of the Audio Engineering Society, vol 54, no. 1, 2006

[67] V. Pulkki and T. Hirvonen, *Localization of Virtual Sources in Multichannel Audio Reproduction*, IEEE Transactions on Speech and Audio Processing, vol 13, no. 1, 2005

[68] V. Pulkki, *Compensating displacement of amplitude-panned virtual sources*, Audio Engineering Society 22th Int. Conf. on Virtual, Synthetic and Entertainment Audio pp. 186-195, Espoo, Finland, 2002

[69] V. Pulkki, *Spatial Sound Generation and Perception by Amplitude Panning Techniques*, Dissertation of Ville Pulkki, which consists of four JAES-articles and two conference articles, 1997-2001

[70] Sabin, A.T., Macpherson, E.A. & Middlebrooks, J.C., *Human sound localization at near-threshold levels*, Hearing Res., 199:124-34, 2005

[71] Savioja, L., *Modelling Techniques for Virtual Acoustics*, Doctoral Thesis, Helsinki University of Technology, Telecommunications Software and Multimedia Laboratory, Report TML-A3, 1999

[72] Scarpaci, J. W. (2006). "Creation of a System for Real-Time Virtual Auditory Space and its Application to Dynamic Sound Localization," Biomedical Engineering. Boston, Boston University

[73] Scarpaci, J. W., Colburn, H. S., and White, J. A., *A System for Real-Time Virtual Auditory Space (PREPRINT*), Proceedings of ICAD 05-Eleventh Meeting of the International Conference on Auditory Display, 2005a

[74] Scarpaci, J. W. and Colburn, H. S., *Principal Components Analysis Interpolation of HRTF's Using Locally Chosen Basis Functions,* J. Acoust. Soc. Am.117, 2561, 2005b

[75] Scarpaci, J. W. and Colburn, H. S., *A real-time virtual auditory system for spatially dynamic perception research,* J. Acoust. Soc. Am.115, 2599, 2004

[76] Shackleton T.M. and Meddis R. and Hewitt M.J., *The Role of Binaural and Fundamental Frequency Difference Cues in the Identification of Concurrently Presented Vowels*, The Quarterly Journal of Experimental Psychology,47A(3), pp. 545-563, 1994

[77] Sheft S. and Yost W.A., *Binaural modulation detection interference*, J. Acoust. Soc. Am. 102, 1791-1798, 1997

[78] Ville P. Sivonen and Wolfgang Ellermeier, *Effects of direction on loudness for wideband and reverberant sounds*, To be presented at the 120th AES Convention, May 20-23, Paris, France, 2006

[79] T. Takeuchi, P.A. Nelson, O. Kirkeby and H. Hamada, *Influence of Individual Head Related Transfer Function on the Performance of Virtual Acoustic Imaging Systems*, presented at the104nd AES Convention, Amsterdam, Netherlands. AES preprint 4700 (P4-3) 16-19 May, 1997a

[80] T. Takeuchi, P. A. Nelson, O. Kirkeby and H. Hamada, *The Effects of Reflections on the Performance of Virtual Acoustic Imaging Systems*, pp. 955-966, in Proceedings of the Active 97, The international symposium on active control of sound and vibration, Budapest, Hungary, August 21-23, 1997b

[81] M. Tikander, A. Härmä, and M. Karjalainen, "*Acoustic positioning and head tracking based on binaural signals*, 116th AES Convention, preprint 264, Berlin, Germany, 8 - 11 May 2004.

[82] N. Tsingos and O. Warusfel, *Perception spatiale du son*, Traite de la réalité Virtuelle. volume "L'homme et l'environnement virtuel", ed. P. Fuchs, G.Moreau (Les Presses de l'Ecole des Mines de Paris, Paris), 2006

[83] O. Warusfel and I. Viaud-Delmon and O. Delerue, *Binaural rendering assessment in the context of augmented reality* , CFA-DAGA, Strasbourg, 2004

[84] Wenzel, E. M., *Localization in virtual acoustic displays*, Presence, 1, 80-107, 1992

[85] E. M. Wenzel (2003) Effect of increasing system latency on localization of virtual sounds with short and long duration, [Invited talk] Workshop on Spatial Media, Aizu-Wakamatsu, JAPAN, March 6-7

[86] Wightman, F. L., & Kistler, D. J., *Headphone simulation of free-field listening*, I: Stimulus synthesis. Journal of the Acoustical Society of America, 85, 858-867, 1989

[87] Wightman, F. L., Kistler, D. J., Foster, S. H., Abel, J. , *A comparison of head- related transfer functions measured deep in the ear canal and at the ear canal entrance*, Abstracts of the 17th Midwinter Meeting, Association for Research in Otolaryngology, 71, 1995

[88] Winn, W.D. (2005). What We Have Learned About VR and Learning and What We Still Need to Study. In Proceedings of Laval Virtual 2005.

[89] Yost William A., *Tone-on-Tone Binaural Masking*, Indiana Mathematical Psychology Programme, Indiana University, Report Number 10-70, 1970

[90] Yost William A., *Weber's Fraction for the Intensity of Pure Tones Presented Binaurally*,Perception and Psychophysics, 11, (1A), 61-64, 1972

[91] Yost William A., *Comments on 'Lateralization and Binaural Masking-Level Difference'*, (G.B. Henning, J. Acoust. Soc. Am., 57, 1975), Journal of the Acoustical Society of America, 57, 1214-1216, 1975

[92] Yost William A. and Dye Raymond*, Binaural Psychophysics, Seminars in Hearing: Binaural Issues in Clinical and Rehabilitative Audiology*, Thieme Medical Publishers, Inc. Vol 18, Num. 4, p321-344, 1997

[93] Zahorik P. and Wightman F. L., *Loudness constancy with varying sound distance*, Nature Neuroscience, 4, 78-83, 2001

[94] F. Zotter, M. Noisternig, *Near- and Farfield Beamforming using Spherical Loudspeaker Arrays*, Proc. Int. Congress of the Alps Adria Acoustics Association, Graz, Austria, September 27-28, 2007

290

Appendix B

First and Second Order Ambisonic Encoding and Decoding equations

Loudspeaker (Az El R)	Loudspeaker (x y z)	W	X	Y	Z
22.5° 0° 1m	0.9239 0.3827 0.0000	0.1768	0.2310	0.0957	0.0000
67.5° 0° 1m	0.3827 0.9239 0.0000	0.1768	0.0957	0.2310	0.0000
112.5° 0° 1m	-0.3827 0.9239 0.0000	0.1768	-0.0957	0.2310	0.0000
157.5° 0° 1m	-0.9239 0.3827 0.0000	0.1768	-0.2310	0.0957	0.0000
202.5° 0° 1m	-0.9239 -0.3827 0.0000	0.1768	-0.2310	-0.0957	0.0000
247.5° 0° 1m	-0.3827 -0.9239 0.0000	0.1768	-0.0957	-0.2310	0.0000
292.5° 0° 1m	0.3827 -0.9239 0.0000	0.1768	0.0957	-0.2310	0.0000
337.5° 0° 1m	0.9239 -0.3827 0.0000	0.1768	0.2310	-0.0957	0.0000

Table 1. The coefficients for the decoding of 1st Order Ambisonic into an Octagon 2D loudspeakers' array. The negative values correspond obviously to a phase inversion (180°).

Loudspeaker (Az El R)	Loudspeaker (x y z)	W	X	Y	Z
45° -35.25° 1m	0.5774 0.5774 -0.5774	0.1768	0.2165	0.2165	-0.2165
315° -35.25° 1m	0.5774 -0.5774 -0.5774	0.1768	0.2165	-0.2165	-0.2165
225° -35.25° 1m	-0.5774 -0.5774 -0.5774	0.1768	-0.2165	-0.2165	-0.2165
135° -35.25° 1m	-0.5774 0.5774 -0.5774	0.1768	-0.2165	0.2165	-0.2165
45° 35.25° 1m	0.5774 0.5774 0.5774	0.1768	0.2165	0.2165	0.2165
315° 35.25° 1m	0.5774 -0.5774 0.5774	0.1768	0.2165	-0.2165	0.2165
225° 35.25° 1m	-0.5774 -0.5774 0.5774	0.1768	-0.2165	-0.2165	0.2165
135° 35.25° 1m	-0.5774 0.5774 0.5774	0.1768	-0.2165	0.2165	0.2165

Table 2. The coefficients for the decoding of 1st Order Ambisonic into a Cube 3D loudspeakers' array. The negative values correspond obviously to a phase inversion (180°).

Loudspeaker (Az El R)	Loudspeaker (x y z)	W	X	Y	Z
0° 90° 1m	0.0000 0.0000 1.0000	0.1768	0.0000	0.0000	0.2500
0° -90° 1m	0.0000 0.0000 -1.0000	0.1768	0.0000	0.0000	-0.2500
40° 26.55° 1m	0.7236 0.5257 0.4472	0.1768	0.1809	0.1314	0.1118
216° -26.55° 1m	-0.7236 -0.5257 -0.4472	0.1768	-0.1809	-0.1314	-0.1118
320° 26.55° 1m	0.7236 -0.5257 0.4472	0.1768	0.1809	-0.1314	0.1118
144° -26.55° 1m	-0.7236 0.5257 -0.4472	0.1768	-0.1809	0.1314	-0.1118
108° 26.55° 1m	-0.2764 0.8507 0.4472	0.1768	-0.0691	0.2127	0.1118
288° -26.55° 1m	0.2764 -0.8507 -0.4472	0.1768	0.0691	-0.2127	-0.1118
252° 26.55° 1m	-0.2764 -0.8507 0.4472	0.1768	-0.0691	-0.2127	0.1118
72° -26.55° 1m	0.2764 0.8507 -0.4472	0.1768	0.0691	0.2127	-0.1118
180° 26.55° 1m	-0.8944 0.0000 0.4472	0.1768	-0.2236	0.0000	0.1118
0° -26.55° 1m	0.8944 0.0000 -0.4472	0.1768	0.2236	0.0000	-0.1118

Table 3. The coefficients for the decoding of 1st Order Ambisonic into a Dodecahedron 3D loudspeakers' array. The negative values correspond obviously to a phase inversion (180°).

Loudspeaker (Az El R)	Loudspeaker (x y z)	R	S	T	U	V
22.5° 0° 1m	0.9239 0.3827 0.0000	0.0000	0.0000	0.0000	0.1768	0.1768
67.5° 0° 1m	0.3827 0.9239 0.0000	0.0000	0.0000	0.0000	-0.1768	0.1768
112.5° 0° 1m	-0.3827 0.9239 0.0000	0.0000	0.0000	0.0000	-0.1768	-0.1768
157.5° 0° 1m	-0.9239 0.3827 0.0000	0.0000	0.0000	0.0000	0.1768	-0.1768
202.5° 0° 1m	-0.9239 -0.3827 0.0000	0.0000	0.0000	0.0000	0.1768	0.1768
247.5° 0° 1m	-0.3827 -0.9239 0.0000	0.0000	0.0000	0.0000	-0.1768	0.1768
292.5° 0° 1m	0.3827 -0.9239 0.0000	0.0000	0.0000	0.0000	-0.1768	-0.1768
337.5° 0° 1m	0.9239 -0.3827 0.0000	0.0000	0.0000	0.0000	0.1768	-0.1768

Table 4. The coefficients for the decoding of 2nd Order Ambisonic into an Octagon 2D loudspeakers' array. The negative values correspond obviously to a phase inversion (180°).

Loudspeaker (Az El R)	Loudspeaker (x y z)	R	S	T	U	V
45° -35.25° 1m	0.5774 0.5774 -0.5774	0.00	-0.1875	-0.1875	0.00	0.1875
315° -35.25° 1m	0.5774 -0.5774 -0.5774	0.00	-0.1875	0.1875	0.00	-0.1875
225° -35.25° 1m	-0.5774 -0.5774 -0.5774	0.00	0.1875	0.1875	0.00	0.1875
135° -35.25° 1m	-0.5774 0.5774 -0.5774	0.00	0.1875	-0.1875	0.00	-0.1875
45° 35.25° 1m	0.5774 0.5774 0.5774	0.00	0.1875	0.1875	0.00	0.1875
315° 35.25° 1m	0.5774 -0.5774 0.5774	0.00	0.1875	-0.1875	0.00	-0.1875
225° 35.25° 1m	-0.5774 -0.5774 0.5774	0.00	-0.1875	-0.1875	0.00	0.1875
135° 35.25° 1m	-0.5774 0.5774 0.5774	0.00	-0.1875	0.1875	0.00	-0.1875

Table 5. The coefficients for the decoding of 2nd Order Ambisonic into a Cube 3D loudspeakers' array. The negative values correspond obviously to a phase inversion (180°).

Loudspeaker (Az El R)	Loudspeaker (x y z)	R	S	T	U	V
0° 90° 1m	0.0000 0.0000 1.0000	0.4167	0.0000	0.0000	0.0000	0.0000
0° -90° 1m	0.0000 0.0000 -1.0000	0.4167	0.0000	0.0000	0.0000	0.0000
40° 26.55° 1m	0.7236 0.5257 0.4472	-0.0833	0.2023	0.1469	0.0773	0.2378
216° -26.55° 1m	-0.7236 -0.5257 -0.4472	-0.0833	0.2023	0.1469	0.0773	0.2378
320° 26.55° 1m	0.7236 -0.5257 0.4472	-0.0833	0.2023	-0.1469	0.0773	-0.2378
144° -26.55° 1m	-0.7236 0.5257 -0.4472	-0.0833	0.2023	-0.1469	0.0773	-0.2378
108° 26.55° 1m	-0.2764 0.8507 0.4472	-0.0833	-0.0773	0.2378	-0.2023	-0.1469
288° -26.55° 1m	0.2764 -0.8507 -0.4472	-0.0833	-0.0773	0.2378	-0.2023	-0.1469
252° 26.55° 1m	-0.2764 -0.8507 0.4472	-0.0833	-0.0773	-0.2378	-0.2023	0.1469
72° -26.55° 1m	0.2764 0.8507 -0.4472	-0.0833	-0.0773	-0.2378	-0.2023	0.1469
180° 26.55° 1m	-0.8944 0.0000 0.4472	-0.0833	-0.2500	0.0000	0.2500	0.0000
0° -26.55° 1m	0.8944 0.0000 -0.4472	-0.0833	-0.2500	0.0000	0.2500	0.0000

Table 6. The coefficients for the decoding of 2nd Order Ambisonic into a Dodeca-hedron 3D loudspeakers' array. The negative values correspond obviously to a phase inversion (180°).

Appendix C

Recapitulative table on the virtual surround, binaural and transaural techniques and systems in the consumer market

Legend:

For every cell

- Unkn: information was not available
- NotClear: information was present, but not particularly clear in terms of description of how the actual technique works.

System type

- Headph: Headphones system with processor
- Box: Hardware processing unit
- Loudsp: Loudspeakers system with processor
- Alg: Algorithm implemented in different hardware and software systems
- Plugin: Audio processing plugin
- StandAlone: Standalone application
- Library(C, C++, MaxMSP…): Library (the programming language is written between the brackets) for audio processing.

Spatialization technique

- Conv(HRIR name): Convolution with HRIR; between the brackets there is, if known, the HRIR database used
- EqDelay: Equalization filter and delay line
- ITD-ILD: Simulation of the Interaural Differences
- StEnhHeadph: Stereo enhancement over headphones based on various phasing effects
- StEnhSpk: Stereo enhancement over loudspeakers based on various phasing effects
- MulDrPh: Multiple drivers loudspeakers working on phase shifts
- MulDrRefl: Multiple drivers loudspeakers working on reflections from walls
- StSpkTrans: Standard speakers with transaural crosstalk-cancellation technique
- StSpkPh: Standard speakers working on phase shifts and other unknown processing.

Functions

- StEnh: Stereo enhancement
- SourPos: Source positioning
- Trans: Transaural function with cross-talk cancellation
- SurTOHeadph: Virtual surround with headphones output
- SurTOSpk: Virtual surround with loudspeakers output
- EnvSim: Environmental simulation
- HeadTr: Head tracking functions
- HRIRSel: Selection of HRIR from different database.

Environmental Simulation

- StRev: Standard stereo reverb
- SurRev: Surround reverb
- BRIR: Binaural Room Impulse Response

Name	System Type	Spatialization Technique	Functions	Environmental Simulation	Additional Info	Commercially Available Products
AKG IVA	Headph Box	ITD-ILD	StEnh SurTOHeadph	-	Conversion from Logic 7 and Dolby Pro Logic to binaural	EARO 777 Quadra EARO 999 Audiosphere II BAP 1000
Beyerdynamic Binaural Environment Modelling	Headph	Conv(various HRIR)	SurTOHeadph EnvSim HeadTr	SurRev	Conversion from discrete multichannel (5.1) to binaural	Headzone Headzone PRO Headzone PRO XT
Bose Articulated Array® Speaker Design	Loudsp	StSpkPh	SurTOSpk	-	Multichannel rendering through two frontal loudspeakers	Bose 1-2-3 Series II DVD Home Entertainment System
Creative Labs CMSS-3D	Alg Headph	UnKn	SurTOHeadph	-	EAX programming interface for gaming applications	X-Fi audio cards HQ 2300 D (implementing Dolby Headphones)
Crossfeed Plugins	Alg	StEnhHeadph	StEnh	-	Plugins family (from different companies) for the simulation of loudspeakers over headphones	Wavelab Externalizer VNOPhones Canz3D Crossfeed Eq
CyberTeam LTD Binaural Mixer	StandAlone	StEnhHeadph	StEnh	-	Based on the theory of binaural beats	Binaural Mixer
Dolby Virtual Speaker	Alg	EqDelay StSpkTrans	SurTOSpk EnvSim	Unkn	Rendering of Dolby Pro Logic II and Dolby Digital through two frontal loudspeakers	Denon S-301 JVC EX-A1

Dolby Headphones	Alg	UnKn	SurTOHeadph EnvSim	Unkn	Conversion from Dolby Pro Logic II and Dolby Digital to binaural	Sony MDR-DS3000 Sony MDR-DS8000 JVC Surround Headphone Adaptor
Holistiks Amphiotik Technology	Alg StandAlone	Conv NotClear	StEnh SourPos Trans	BRIR NotClear	Two different software families for binaural source positioning and stereo enhancement.	Amphiotic Synthesis Amphiotic Enhancer PR, ST and LT
KEF KIT (Kef Instant Theatre)	Loudsp	MulDrRefl	SurTOSpk StEnh	-	Two speakers with multiple drivers oriented in different directions	KEF KIT
NIRO SIP (Sur-round Image Processor)	Loudsp	MulDrPh	SurTOSpk StEnh	-	Unique central component with multiple drivers oriented in different directions	Niro 1.1 Pro II
Pioneer Direct Diffuse	Loudsp	MulDrRefl	SurTOSpk StEnh	-	Two speakers with multiple drivers oriented in different directions	HTP 3600 SE-DIR800C Cordless Surround (implementing Dolby Headphones)
PolkAudio SDA Surround Bar	Loudsp	MulDrRefl	SurTOSpk StEnh	-	Unique central component with multiple drivers oriented in different directions	SDA Surround Bar

QSound Labs	Alg StandAlone Plugin Headph	Conv EqDelay NotClear	StEnh SourPos SurTOHeadph SurTOSpk Trans	-	Different standalone modules and audio plugins with different functions for binaural and transaural output	Acoustic Research Model AW791 UltraQ - QMAX II iQms2 QSYS QX Q123 QCreator QVE
Sensaura	Alg StandAlone	Conv(Kemar and others)	StEnh SourPos SurTOHeadph SurTOSpk Trans EnvSim HRIRSel	UnKn	VirtualEar technology for the selection and customization of HRTF	3DPA macroFX-zoomFX-environmentFX STC Multidrive VirtualEar Game CODA Headphone Theater Algorithm
Sony S-Force PRO	Loudsp	MulDrPh	SurTOSpk StEnh	-	Two standard speakers with a processing unit	BRAVIA Theater System DVD Dream System
SoundHack	StandAlone Plugin	EqDelay	SourPos	-	The binaural filters are derived from the MIT Kemar HRTF (*see* Gardner, 1994)	Soundhack 0.896 +binaural
Spatializer Audio Laboratories, Inc	Alg StandAlone Plugin	Conv EqDelay NotClear	StEnh SurTOHeadph SurTOSpk	-	-	Various products offered by OEMs (Original Equipment Manufacturers) licenses and customers, for example Sharp, Apple, Jvc, Acer, Panasonic,

						Toshiba, Hitachi, Sanyo…
Spin Audio 3D – Positional Audio Engine	Alg StandAlone	Conv(different HRIR database)	SourPos	-	The different available HRTF can be tuned and customized	3DPanner Studio
SRS Tru-Surround	Alg StandAlone	Conv EqDelay NotClear	StEnh SurTOSpk Trans	-	-	Various systems from Marantz, Sony, Tannoy, etc…
SRS Headphones	Alg StandAlone	EqDelay	StEnh	-	-	Sennheiser RS 130
Toltec	Alg Headph	EqDelay	StEnh SurTOHeadph	-	Conversion from Dolby Pro Logic to binaural	Sennheiser HD 580 Cyclone 3D (from Vitual Listening System, Inc)
TC Electronic System 6000	Alg Box	Conv(different HRIR database) EqDelay	SurTOHeadph EnvSim	SurRev	Complete unit for multichannel audio processing	System 6000
Wavearts Panora-ma	StandAlone Plugin	Conv(MIT Kemar and CIPIC) EqDelay	SourPos Trans EnvSim	UnKn	For the HRTF sets, *see* Gardner, 1994, and Algazi, 2001	Wavearts Panorama (standalone or plugin version)
Yamaha Silent Cinema and Virtual Cinema DSP	Alg Box Loudspk	EqDelay StSpkTrans	StEnh SurTOHeadph Trans	-	-	DVX series (S203, S301, S200, S150 and S30)
Yamaha Digital Sound Projector	Loudspk	MulDrRefl	StEnh SurTOSpk	-	Unique central component with multiple drivers	YSP series (1000, 800 and 1)

300

Appendix D

Table on the virtual surround, binaural and transaural techniques and systems in the consumer and professional market, with quality evaluation

Name	General Information	Spatialization Technique	Functions Summary	Quality Evaluation	Related Products
AKG IVA *www.akg.com* *Available for the Consumer Market*	IVA uses advanced signal processing to reproduce the amplitude and phase shifts that our ears and brain uses to localize sound sources. The first objective of IVA is to move the apparent sound source outside the head, giving to the listener the perception of real sound sources located in any position around him/her. IVA is perfectly integrated with LOGIC7: this technology from Lexicon emphasizes the stereophonic signal using localization cues (simulations of HRTF through equalization filters and delay lines) to envelop the listener in a three-dimensional soundscape.	Simulation of interaural phase and level differences. It seems that no simulation is performed for the Direction Dependent Filtering.	IVA+LOGIC7 (conversion from 8 channel surround to binaural stereo) IVA+Dolby Pro Logic (conversion from Dolby surround formats to binaural stereo) IVA+Stereo (simulation of a standard stereo setup)	First of all, the demo is made in Flash, and the sound signal has obviously been highly compressed with lossy algorithms. For this reason, comb filtering is particularly noticeable for high frequencies; this has, of course, in terms of realism a negative effect on the performance of the spatialization. When what is claimed to be a standard stereo voice (and is actually a simple mono signal) is compared to the virtual surround simulation, processing for the enhancement of the space perception (as are stereo spread algorithms based on frequency panning) is clearly audible: the virtual surround signal is much louder than the mono, and with much more energy on the low frequencies, a fact that can be especially useful in terms of binaural spatialization for the ITD perception. This obviously acts as an advantage for the spatial perception of the virtual surround: a more “fair” comparison, using for example the same signal with the same	AKG HEARO 777 Quadra AKG HEARO 999 Audiosphere II AKG BAP 1000

				level and frequency response, would have helped in understanding better the actual features of the algorithm. The frontal localization is not particularly effective, and the impression is of a sound source located in the centre of the head. The rear localization works definitively better: the sound seems to emerge from the head, and the sensation of a source located in the back is quite effective. Nevertheless, the sound source position is not at all clear and localizable. It seems they are using the simulation widely for the ITD cue, also because the algorithm is particularly efficient for low frequency sounds: still, a good simulation of the DDF does not seem to be performed, and the ILD seems to be used as a frequency independent cue (*see* Chapter 3). The overall evaluation is not particularly positive; the binaural spatialization algorithm seems too simple, far too oriented towards the enhancement of bass frequencies and the indiscriminate creation of unreal and over-enhanced auditory spatial fields.	
Beyerdynamic Binaural Environment	It is a family of algorithms for the binaural spatialization with	Little information is given about how	Conversion from	Listening to 5.1 surround musical recording and mixing, the sensation of auditory envelopment	Beyerdynamic

Modelling *www.beyerdynamic.com* *Available for the Consumer Market*	environmental modelling that is mainly used for conversion between surround sound formats (limited to 5.1) and binaural. There are not many information about the processing performed to obtain the binaural rendering, with room simulation, of the surround channels: the most probable hypothesis is that all the environmental simulations are carried out on a 5.1 format, then every single channel is converted to binaural using a virtual loudspeakers system (*see* Chapter 7). The real novelty of the systems based on this technique, compared to other consumer available binaural systems, is the fact that a head tracking system is implemented, so that the sound-field can rotate following the position of the head of the listener, which is tracked using two ultrasonic transmitters placed above the headphone. So far, only the azimuth angle is tracked, therefore the sound-field can be rotated only in the horizontal	the sound spatialization and the environmental simulation are performed. It seems that the conversion between multichannel and binaural is performed through convolution with HRIR using a virtual loudspeaker system (*see* Chapter 7), and that the environmental simulation is calculated in the multichannel domain then converted to binaural using the same technique. It uses different HRTFs, giving the user the opportunity to import his/her own.	discrete multichannel (5.1) to binaural stereo, with environmental simulation and head-tracking.	is certainly effective. Comparing a standard stereo rendering with the binaural simulation obviously helps in localizing sound sources outside the head and in gaining the impression of being surrounded by them. The environmental acoustic simulation seems to help widen the surround sound image, but unfortunately there are only two available controls on the simulated room acoustics, and these are not related to any familiar room parameter, such as reverb time, low and high damping and diffuseness. The head tracking definitely helps in gaining a more realistic impression of a proper 3D sound-field, as well as in helping for the solution of front/back confusions. However, the sound-field rotations seem to be enhanced, at least for the frontal sound sources: a movement of the head of 30° of azimuth seems to correspond to a rotation of 60° of the frontal sound sources. Trying to express an overall evaluation of this system, it is possible to say that the Beyerdynamic binaural technique clearly represents a major step forward from a standard stereo listening: the virtual surround image is wide; thanks to the environmental simulation and to the head-tracking, the sound sources seem to be located outside the head of the	Headzone Beyerdynamic Headzone PRO Beyerdynamic Headzone PRO XT

	plane.			listener. Nevertheless, the localization of distinct sound sources is not particularly accurate, and the head-tracking does not seem to be realistically implemented.	
Dolby *www.dolby.com* *Available for the Consumer Market*	**Dolby Virtual Speaker** A virtual surround system that can be used paired with Dolby Pro Logic II and with Dolby Digital surround sound formats. It works through simulations of environmental acoustics, trying to reproduce the sound spectrum and dynamics typical of a 5.1 system. It implements crosstalk cancellation and room simulation techniques, based above all on the reproduction of reflections from the walls of the room and on the complex management of the attack and sustain of sounds. **Dolby Headphones** It is a binaural algorithm for	For neither system was it possible to gather enough information about the spatialization technique. Dolby Virtual Speaker seems to work on cross-cancellation techniques coupled with environmental simulation. Dolby Headphones works by performing a conversion between discrete multichannel (5.1) and binaural stereo using a virtual loudspeaker system (*see* Chapter 7).	Conversion from discrete multichannel (5.1) to binaural stereo.	ONLY FOR DOLBY HEADPHONES The demo files (downloadable from the Dolby site) are not particularly impressive: the sound image certainly becomes wider in the "virtual surround" mode, yet this seems to be achieved only through simple environmental simulation and interaural delays ratherthan through a proper binaural simulation.	Dolby Virtual Speaker Denon S-301 JVC EX-A1 Dolby Headphones; Sony MDR DS3000 and DS8000 JVC Surround Headphone Adaptor

	surround through headphones reproduction. Analyzing the material given in the Dolby internet site, it seems that this coding is based on the binaural simulation of the five sound sources (Dolby Digital 5.1). However, it is not specified whether this simulation is done through convolution with HRIR or equalization filtering plus delay lines.				
Holistiks Engineering Systems Amphiotik Technology *www.holistiks.com* *Available for the Consumer Market*	It is a binaural algorithm implemented in a series of processors for the binaural spatialization using HRTF simulation through convolution with HRIR and environmental simulation using BRIR (Binaural Room Impulse Response). Within the collection of processors, there is also an implementation of a transaural engine, with cross-talk cancellation features.	It is not really clear whether the spatialization is performed using a proper HRIR convolution of performing simplified operations such as limited frequency filtering and delay lines.	Amphiotik Synthe-sis Audio mixing and mastering system with Binaural Synthesis. It includes transaural audio (cross-talk cancellation), reflection modelling and reverberation. Amphiotik Enhan-cer 3D-Audio system that allows to place the channels of a	In the Holistic internet site it is possible to download a demo version of the Amphiotik software. The PR version has been tried using Winamp as the host platform, and the following considerations have been performed. The interface is certainly interesting, and it allows the control of a large number of parameters: the choice between binaural and transaural listening, loudspeaker positioning in the space, post equalization filtering options, Automatic Gain Control (something like a compressor-limiter), SFM (Sound Field Model) with dry/wet controller, VW (Virtual Word), with the possibility to choose the type of room to be simulated (anechoic, medium, large, etc.), and a choice among the materials of the	Amphiotic Synthesis Amphiotic Enhancer PR, ST and LT

			track in a fully customizable 3D Virtual Auditory Environment, that is the full implementation of the Amphiotik Technology. Enhancer LT (Lite Edition) A 3D-Audio system that converts simple stereo formats to an enhanced “3D stereo output”. It allows also the placing of virtual sound sources in a 3D space, but this can be achieved merely by using presets (not customizable).	reflecting-absorbing surfaces of the different rooms The quality of the processing does not seem to be as good as the interface appears: the sound seems to be surrounding the listener, but the localization of sound sources is far from being clear, and different artificial colourations, such as comb filtering and chorus, seem to be present. The rear sound sources are not clearly audible, and most of the sounds seem to come directly from inside the head (IHL). The reverb is possibly one of the weakest parts of the whole algorithm. Without considering the signal colourations cited above, the quality of the simulation is far from being perfect, and it is obviously closer to a standard stereo reverberation than to a proper 3D environmental simulation. The cross-talk cancellation for transaural listening does not seem to work properly. Even when the head is positioned exactly central in front of the stereo loudspeakers system, the 3D sound effect is far from being satisfactory, and the impression is similar to what could happen when the signals at the two loudspeakers are out of phase. The amount of controllable options and the fact	

				that these software processors are very cheap make them an interesting option if something more than stereo reproduction needs to be achieved, yet this is far from being a proper binaural simulation.	
QSound Labs *www.qsound.com* *Available for the Consumer Market*	QSound Labs has been doing research into virtual surround technologies since 1990. In the past four years, they have implemented technologies for the virtual surround in monaural (Q123 technology) and binaural (QXpander technology, for both binaural and transaural), and for discrete surround (QSurround, for both stereo and multichannel systems).	All of the QSound techniques use a simulation of the HRTF directly derived from the study of the HRIR measured from a dummy head; the simulation is done, for the three localization cues, through equalization filtering and delay lines.	Positional 3D Input through a mono or stereo signal, it can simulate in every position a virtual sound source for a two-channel (virtual surround) or a multichannel (discrete surround) listening. It can also be input with more signals (the main asset of this algorithm is the ability to mix binaurally spatialized sounds together). The virtual surround stage works on a HRTF simulation	On the QSound internet site it is possible to listen to and to download some demo software. Here follow some considerations about 3D audio demos for binaural and transaural listening: • The sensation is similar to that perceived when there are phasing problems between the two loudspeakers: sound sources are definitely not clearly localizable. The movements are not well implemented: the sound sources seem to move from one loudspeaker to the other with no interpolation. The elevation parameter seems randomized: sometimes it is possible to hear a sound coming from the vertical plane, yet it really sounds as a random processing. • The comparison between 3D Stereo and standard stereo is quite useful: in this way it is possible to understand that there are differences between the two, even if it is not clearly possible to identify these. Nevertheless, it seems as though the spati-	Acoustic Research Model AW791 A standalone headphones system for the conversion from Dolby Digital 5.1 format to binaural (QMMS) UltraQ - QMAX II Stereo to binaural converter (hardware and software, based on the QXpander algorithm) iQms2 A plugin version of QMAX II for

			technique, based on equalization filtering and delay lines, with crosstalk cancellation techniques for the transaural listening. <u>QXpander</u> This technology is similar to AKG IVA. Input through a stereophonic signal, it emphasizes its spatial characteristics, eliminating the IHL effect using interaural phase shifts and reverb. <u>Q123</u> It combines QXpander and its stereophonic expansion with a mono input stage; the role of this algorithm is to create and "spread" a stereo image from	alization is performed through level and time differences, and various phase shifts: these elements, which do not seem to follow a proper simulation of the HRTF, make the spatialization effect slightly randomized, and sound sources cannot be properly localized. As a final consideration, it is possible to say that with 3D Stereo from QSound the perception of a more "surrounding" soundscape is achieved, although this is far from a real surround sensation with clearly localizable sound sources.	Windows Media Player, RealONE Player and WinDVD Player. Qtools/AX A collection of three plugins in Direct-X format. QSYS An implementation of Positional 3D. QX An implementation of QXpander. Q123 An implementation of Q123. Qcreator An audio editor that allows the creation of a 3D effect on every audio track; the main function is the possibility to

			a mono input, trying to help the listener to localize sound sources outside the head. <u>Multi-Speaker (QMSS)</u> It creates 2.1, 3.1, 4.1, 5.1, and 7.1 channel renderings from a mono or stereo input. <u>QSurround</u> It converts multichannel formats, as Dolby Digital, into binaural for a headphone or loudspeaker listening (binaural or transaural). It implements a HRTF simulation tech-nique based on equalization filtering and delay lines. The transaural stage implements a		automate the movements of a virtual sound source in the space. QVE (Qsound Virtual Engine) A digital audio engine that can be embedded in consumer products (like Philips Consumer Electronics, various sound-cards or software for PC) which support the implementation of the QSound plugins. MicroQ QXpander and QVE implementa-tion for mobile phones (Symbian OS).

			crosstalk cancellation technique. It can also encode two channels into surround format, and it can emphasize the surround perception for multichannel formats.		
Sensaura *www.sensaura.net* *Available for the Consumer Market*	The Sensaura technology, called Virtual Ear™, uses an HRTF library set representative of the population mean. The user can change different parameters, related to the shape and dimensions of the ear and of the head of the listener, in order to select and adjust the HRTF. For the achievement of the optimal HRTF, the user can modify: • Ear size and shape: the measurement of the HRTF database has been carried out after multiple studies on the different typologies of external ears (and of head dimensions): the use of an	This is the first commercially available binaural algorithm family to use HRTF with convolution between the input signal and a HRIR database, and that allows the user to select and modify the HRTF depending on their own ear physiology and head dimensions. Convolution with different HRIR	3DPA (3D Positional Audio) A 3D positional audio algorithm, the "hearth" of Sensaura technology. Digital Ear HRTF library the HRTF database. MacroFX Algorithm for the binaural spatialization for closely positioned sound sources.	On the Sensaura internet site it is possible try demos of the Sensaura software. Here follow some considerations about the Player 3D demo: • The plugin has few controls (and, in the demo version, many of these are disabled): two kinds of HRTF (lite and full), azimuth, elevation and distance (disabled), reverb (it was disabled, and it was not possible to see which parameters could have been used), and type of listening (stereo standard, binaural, transaural, quadraphonic or 5.1). • The transaural listening seems to give good spatial perception: the back position seems to be reached, but the sound source movements are not particularly smooth, and it is clearly possible to perceive strange phase shifts passing from the left to the right	All the algorithms cited on the third column (Functions Summary) have their respective software implementation (using the same name) Game CODA Platform for the computer games developers. It helps developers to integrate Sensaura sound algorithm in their

1 *See* http://www.knowles.com

	artificial head (KEMAR, Knowles Electronic Manikin for Acoustic Research[1]) with four different ear types (DB-060/DB-061—DB-065/DB-066—DB-090/DB-091—DB-095/DB-096 from KEMAR) has been exploited, as has the development of the Sensaura Digital Ear technology that helped in the creation of a certain number of HRTFs included in the Sensaura Library. • Head-Related Transfer Function: the HRTF is measured in an anechoic chamber, at one metre from the listener, and is then separated into three factors: o Far-ear response o Near-ear response o Interaural time delay. • Inter-Aural Time Delays (ITD): they strongly depend on the size of the head, therefore they need to be scaled depending on the user's physiology. Sensaura	database, chosen dependently on the ear shape and head circumference of the listener, is performed. Regarding the environmental simulator, no information was available about which techniques are used.	<u>ZoomFX</u> Algorithm for the re-creation of the acoustic size of sound-emitting objects. <u>EnvironmentFX</u> Technology for the emulation of a range of acoustic environments; it can be used as an effect, or to help the localization of sound sources outside the head. <u>XTC</u> Crosstalk cancellation technique for the virtual surround (transaural). <u>Multidrive</u> Virtual spatialization through loudspeakers for computer games	hemisphere, and *vice versa*. • The binaural listening does not seem to work as well as the transaural one: the back position seems to be reached, but the filtering effect sounds very bad. It is impossible to perceive a frontal sound source, and the phase shifting, passing from the left to the right hemisphere, is particularly annoying. The localization of apparent sources outside the head is definitely not achieved. • It seems to be really lightweight in terms of computational elaboration. About the binaural recording (done by Sensaura with a dummy head): • Not particularly impressive: the sound sources all seem to be located inside the head, and it is difficult to locate frontal and back sound sources. The overall judgement on the demos is not particularly positive. It could be useful to test the Player 3D algorithm again with the reverb activated (as said before, it was disabled in the demo version).	software (game, design, music…). Headphone Theater Algorithm for the conversion from 5.1 surround to binaural.

	uses a mathematical model to scale the ITD for all different head types. • Virtual Ear technology: it enables users to select from a library an HRTF corresponding to their own physical dimensions. The most important feature of this technique is the ability to scale the ear dimensions and head dimensions independently, as a duplex system, and then to combine the results in order to provide a wide range of physiological permutations from which to choose. This technology is based on the concept that the complex resonant and diffractive effects integral to the HRTF can be independently scalable. The following parameters can be modified by the user in order to select or create the right HRTF: o HEAD SIZE o EAR SIZE o CONCHA DEPTH		<u>Virtual Ear</u> CAD model of the human ear. It is a software for the customization of the HRTF filtering.		

	○ CONCHA TYPE The setup procedure for the selection of the proper HRTF is based on several listening stages, as well as on calibrations made from physiological data.				
Spatializer Audio Laboratories, Inc *www.spatializer.com* *Available for the Consumer Market*	For more information about Spatializer Audio Laboratories, Inc, *see* Section 2.1. The Spatializer technology works on psychoacoustical sound localization elements for creating a surround sound effect via stereo frontal loudspeakers or headphones. Spatializer patented and proprietary technology focuses on the use of psychoacoustics, the study of how sound waves are perceived by the human ear, and neural science, the study of how perception is conceptual-ized in our minds. The various algorithms in which this technique is implemented work on the physiological aspect of the human hearing, simulating a virtual sound environment through the study of the localization cues.	The spatialization is done through a simulation of the HRTF: it is not clear whether this is performed using a proper HRIR convolution of performing simplified operations such as equalization filtering and delay lines.	Spatializer N-2-2 Ultra™ Virtual simulation of a surround sound field through two loudspeakers (inputted with a surround sound signal). Spatializer Virtual-Surround VBX™ Surround sound enhancement, being inputted with just two channels. Spatializer Vir-tuaLFE™ Psychoacoustic simulation of the subwoofer channel. Spatializer PCE™	On the Spatializer internet site it is possible to download a demo version of the VSP-11 (Virtual Surround Processor) plugin for PC (unfortunately, it was impossible to download demo files about the VBX). Here follow some brief considerations about it: • On the interface, it looks like a game • The 3D control knob (it should be a control of the depth of the surround sensation) does not seem to work in terms of binaural lis-tening; the transaural setup seems to work as a stereo spreader: the stereo image is larger, but back sound sources could neither be heard nor properly localized. The overall evaluation of the demo cannot be considered positive: the quality of the 3D enhancer spatialization is not at all impressive, but this may be linked to the fact that it is only a demo version, with less actively controllable parameters. It would be useful to listen to signals spatialized using the VBX algorithm.	Spatializer Audio Laboratories, Inc technologies are incorporated in products offered by OEMs (Original Equipment Manufacturers) licenses and customers, for example Sharp, Apple, Jvc, Acer, Panasonic, Toshiba, Hitachi, Sanyo… VSP-11 Surround and virtual surround sound plugin for windows

			Simply a digital filter implemented in order to improve the quality of digital audio, for example, through the internet. Spatializer Vi.B.E™ bass enhancement for small loud-speakers. Spatializer Natural Headphone: virtual surround through headphones. It works both on a simulation of the HRTF and on room reverberation modelling. Spatial-izer((environ))™ Surround sound through two small and close loud-speakers, located either to the front or to the sides of the listener.		

SRS *www.srslabs.com* *Available for the Consumer Market*	General information about SRS can be found in Section 2.1. **SRS Tru-Surround** A binaural algorithm for transaural reproduction. Tru-Surround is a sound-scheme that has the ability to take multi-channel encoded sources, such as Dolby Digital, and reproduce the multi-channel surround effect by just using two speakers **SRS HeadphoneTM** A binaural algorithm for headphones reproduction. It works through a simulation of the HRTF, but it is not a "positional" binaural algorithm, in that it does not convert a surround signal into a binaural one, it merely implements a binaural technique to give the impression of sound sources located outside the head of the listener.	Tru-Surround works through a simulation of the HRTF (it is not specified whether this is done with HRIR convolution or simple equalization filtering), with a cross-talk cancellation technique. SRS Headpho-neTM too works on HRTF simulation, but it seems that this is only relative to phase shifts in order to help the listener to localize sound sources outside the head.	SRS Tru-Surround Binaural algorithm for transaural reproduction for the conversion between Dolby Digital multichannel and stereo SRS HeadphoneTM Binaural stereo enhancement algorithm for headphones reproduction	On the SRS internet site it is possible download demo files of the SRS Tru Surround and SRS Headphone. Here follow some considerations about the SRS Tru-Surround: • The demo is the Apollo 13 movie, with a comparison between standard stereo and transaural. • The quality of the sound is quite good, and it is possible to hear a clear spatial en-hancement with the TRU Surround algorithm activated: but this is still far from being a real surround sensation. • Even if the sensation of the space seems enhanced, it is rather difficult to hear clearly and to localize sound sources at the back of the listener. The final judgement about this implementation is quite positive; the stereo sensation really is enhanced, and it is possible to hear lateral sound sources, even if this is quite far from a real 3D transaural listening. It seems as if the stereo enhancement is achieved with phase shifts and with an equalization filtering, but not with a proper HRTF simulation through convolution with HRIR. Here follow some considerations about the SRS Headphones: • The enhancement sensation is similar to that perceived listening to Tru Surround. • Even if the stereo standard is enhanced,	Sennheiser RS130 Implementation of SRS Headpho-neTM

				sound sources still seem rather to be lateralized inside the head than localized outside. The final judgement about this implementation is positive in terms of the stereo enhancement, but the sound sources are still often localized inside the head.	
Wavearts Panorama *www.wavearts.com* *Available for the Consumer Market*	It is a complete binaural (and transaural) spatialization plugin based on HRTF simulation, with HRTF selection functions and environmental simulation.	It implements an HRTF simulation: the Panorama manual states they used a digital filter, in particular an equalizer, to simulate the HRTF. One of the parameters in the plugin's interface is the choice of the HRIR database: the user can choose the CIPIC HRIR database (*see* Algazi, 2001), the MIT HRIR database (*see* Gardner, 1994), or to import his/her own HRIR database. There is	Positional binaural audio processing plugin with HRTF selection and environmental simulation.	On the Wavearts internet site it is possible download a demo version of Panorama. Here follow some considerations about it: • The flexibility is excellent: there are many parameters, and everything seems to work quite well. • The effect of sound sources localized outside the head of the listener is quite good; it is possible to hear that the sound is no more lateralized inside the head, but the sensation is not of a sound source clearly localized outside the head. • The simulation of the distance is quite good yet not excellent; there are problems for the close sound source localization (the manual claims that the distance simulation is done only through the loudness and direct/reflected sound ratio, but it is known that there are also spectral cues for the distances, *see* Chapter 5) • The binaural reverb quality does not seem to be very good: it seems a normal stereo reverb, maybe enhanced with some processing techniques for helping the listener	Wavearts Panorama

		also an implementation of a binaural reverb, with parameters as delay, reflection, size and reverb time.		to localize sound sources outside the head. • About listening through the headphones: it is quite difficult to distinguish the front from the rear. The simulation of the elevation is quite good for the upper hemisphere, but not for the lower. • The transaural listening is not particularly good: sources localized in the rear hemisphere are not at all present, and it seems that there are no differences between the upper and the lower hemispheres. • The simulation of the doppler effect works quite well, both on transaural and on binaural listening • The simulation of the movement of the sound sources is not excellent: it is not possible to gather what HRIR interpolation technique has been implemented, but it seems the differences are only in the ITD and IID localization cues (not DDF). It seems that the algorithm performs a refresh of all the parameters only once each second, therefore the movement of the sound source does not result as smooth. • It is quite difficult to distinguish between the two HRIR databases (MIT and CIPIC) The final judgement is relatively positive: the surround sensation is not excellent yet it is possible to hear a significant difference from standard stereo listening. The fact that the user can import a self-made HRTF and use this for	

				the binaural spatialization is highly positive. Given the information gathered from the internet site and from the Panorama manual, the quality of the implementation of the reverb and distance simulation is rather disappointing.	
CSound hrtfer *Available for the Professional Market*	It is a opcode (a function) for the binaural spatialization within the CSound programming environment.	These unit generators place a mono input signal in a virtual 3D space around the listener by convolving the input with the appropriate HRTF data specified by the opcode's azimuth and elevation values. No information is given about the HRTF data used for the spatialization.	Positional binaural audio processing, with simulation of source movements.	The CSound hrtfer function is freely available over the internet, and it was therefore possible to try it out. The quality of the spatialization is not particularly good: a signal spatialized in a sound source virtually located at 0° of azimuth and 0° of elevation does not sound too different from a normal mono diotic signal. When spatializing the sound sources in different positions on the horizontal plane, the spatial sensation becomes more effective, and the differences between this and the non-spatialized signal are clearly audible. The localization of sound sources behind the head, as well as the simulation of elevation, do not seem to be effective: a source simulated at 45° of azimuth and 45° of elevation sounds nearly the same as a source simulated at the same azimuth degrees but at -45° of elevation, and the same happens for specular sources on the horizontal plane. The simulation of the movement is not particularly smooth (it seems more like a stepped movement); occasionally, strange phasing effects are clearly audible.	CSound hrtfer

				As an overall judgement, the spatialization of the hrtfer function is far from being satisfactory.	
IRCAM Spat *Available for the Professional Market*	It is a complete 3D audio processor implemented as VST plugins, standalone application and MaxMSP object library, from the Spatialization Research Group of IRCAM[2] (Paris, France). Its central features are linked with the management of multiple loudspeaker arrays, while it also has a section for the binaural spatialization.	Spatialization through convolution with HRIR from the Listen HRTF database[3]. The environmental simulation is performed in the multichannel (mainly Ambisonics) domain, then converted to binaural.	Positional binaural audio processing, with simulation of source movements (through HRIR interpolation), Doppler effect and source directivity. Environmental simulation performed in the multichannel domain. Conversion between multichannel formats and binaural.	It has not been possible to test version 4.0 Spat MaxMSP library; nevertheless, an audio quality test has been performed with the 3.0 version. The MaxMSP object library is extremely flexible, allowing the use of different objects with different functions within a chain whose order, links and parameters are decided by the user. Using simply the binaural spatialization engine, with no room simulation, the quality of the spatialization is not particularly good: comparing normal mono signals to spatialized ones, there is the impression of a sound source moving from inside to outside the head, but there certainly exist some confusions for front-back sources. The simulation of the elevation parameter seems to work rather better, and sound sources above are clearly distinguishable from sound sources below the listener. The implementation of the movements of the sound source is very well implemented, as well as the Doppler effect simulation, and sound sources seem to move smoothly around the space surrounding the listener. The environmental simulation clearly helps in	IRCAM Spat 4.0

2 *See* http://www.ircam.fr

3 *See* www.ircam.fr/equipes/salles/listen/

				the localization of apparent sound sources outside the head, but does not seem to be particularly realistic, most of all trying to simulate large halls. The reverb seems not to be particularly directional, even when the simulated room is clearly asymmetrical. The simulation of the directivity of the sound source is very interesting, but not particularly effective: it seems as if the directivity of a sound source (in this case a talking head has been used) is enhanced in order to make the simulation more audible, but this does not seem to reflect the effect in the real situation. As an overall evaluation, the Spat library definitely offers a very interesting and flexible tool for sound spatialization in general. The quality of the binaural spatialization is not particularly good, nor is the quality of environmental simulation, while the simulation of the moving sound source seems to be extremely well done.	
IEM Bin_Ambi *Available for the Professional Market*	IEM Bin_Ambi is a real-time rendering engine for virtual (binaural) sound reproduction, composed by a sophisticate object library for the Pure Data[4] visual programming environment. The majority of the	Spatialization through convolution with HRIR from a HRIR database. The environmental	Positional binaural audio processing. Environmental and source movement simulation performed in the Ambisonics domain.	The IEM Bin_Ambi PD library is freely available under the GNU licence: it is extremely flexible, allowing the use of different objects with different functions within a chain whose order, links and parameters are decided by the user. However, it is not particularly intuitive in terms of its use, and a proficient knowledge of	IEM Bin_Ambi

4 *See* http://www.puredata.org

	library has been programmed by Markus Noisternig and Thomas Musil, within the IEM Sonevir research project.	simulation is performed using a Virtual Ambisonics Approach (*see* Noisternig 2003a and 2003b).	Conversion between High Order Ambisonics and binaural using a Virtual Ambisonics Approach. .	Pure Data is demanded. Generally, the quality of the spatialization is quite positive: the sources seem to be located outside the head (the effect is greater if spatialized sound and a standard mono signals are compared), and the possibility given of rotating the soundfield in the Ambisonics domain using a head-tracking device (in this case, an XSens gyroscope has been used) clearly helps in the correct localization of sound sources. When the environmental simulation is added, the sources seem to move very far away. It seems that the direct-reflected sound ratio is very low, and the simulation does not seem to be realistic. Also, the reverb seems to be more stereophonic than surround: simulating asymmetrical rooms, the asymmetries are not clearly audible, and again the reverb does not seem to change according to the movements of the head. The Virtual Ambisonics Approach is certainly a very flexible and powerful tool for real-time binaural spatialization with environmental simulation. The quality in terms of binaural spatialization using this library seems to be good while the environmental simulation is not satisfactory, and this results in a non-realistic recreation of 3D environments where sound sources seem to be located too far away from the listener.	

Aristotel Digenis Plugins *Available for the Professional Market*	Aristotel Digenis is an Experience Audio Programmer at CodeMasters[5] (an audio games production company), and developed different surround and virtual surround libraries and plugins freely downloadable from his internet site (here, only those related to binaural spatialization are reported): • Bidules: a family of plugins for Ambisonic encoding, decoding, binaural decoding and rotations. • MIT HRTF Library: an open source C library offering access to the MIT Kemar HRTF set through two simple functions.	Spatialization through convolution with HRIR from the MIT Kemar HRTF database (*see* Gardner, 1994).	Conversion between High Order Ambisonics and binaural (it is not specified which approach is followed).	The plugins are freely available on the internet, yet to be used in realtime they need the Plogue Bidule application, which can be downloaded as a demo version. The quality of the spatialization is not particularly good: this is no surprise, because all binaural processors using the MIT HRTF database lack a very high spatialization quality. Nevertheless, a comparison between mono signals and those First Order Ambisonisc binaurally processed gives the clear impression, in the latter, of apparent sources located outside the head. Nevertheless, this depends greatly on the Ambisonics material used. The comparison has been made between anechoically recorded First Order Ambisonics and the same recording performed inside a room; while the former's apparent sound sources seem to remain inside the head, in the latter the impression of a proper realistic soundscape is far clearer. The flexibility of this library, as well as the fact that it is freely available and easy to use (much easier, for example, than the IEM Bin_Ambi PD library), makes it probably the best option for obtaining a binaural conversion from Ambisonics signals. However, using a better HRTF set and performing some 3D environmental situations would certainly have helped increase the perceived spatial quality.	Bidules MIT HRTF C++ Library

5 *See* http://www.codemasters.co.uk

LIMSI LSE *Available internally at LIMSI-CNRS*	This set of objects for MaxMSP is available for neither the consumer nor the professional markets. In fact, it is a library internally developed at LIMSI-CNRS, and for the exclusive use of LIMSI researchers. The author worked at the LIMSI-CNRS lab for one-and-a-half years, and in this period contributed to the development of the library itself. It has therefore been considered of value to report some information about it. The library is of MaxMSP objects for the binaural spatialization of sound signals based on convolution with the IRCAM Listen HRTF database, with an advanced management of ITD (for HRIR interpolation and customization dependently on the head circumference of the user) and ILD (for simulation of close sound sources).	Spatialization through convolution with HRIR from the Listen HRTF database[6]. Advanced ITD and ILD management for the simulation of movements and of close sound sources.	Positional binaural audio processing with simulation of close sound sources (between 1 m and 15 cm)	The library is extremely flexible, giving the user the opportunity to select which object or function to use, and to build his/her own patch specifically for what needs to be achieved (a proficient knowledge of MaxMSP is demanded). The spatialization quality at one metre is sufficiently effective, even if not particularly surprising: the flexibility of the system in terms of HRTF selection and ITD customization allows the best to be gained in terms of spatialization quality with the actual HRTF sets available. The simulation of close sound sources is extremely effective: for example, at 90° of azimuth, it is possible to hear clearly the sound sources approach, up to few centimetres from the head of the listener. It is also possible to programme specific patches for the conversion between High Order Ambisonics and binaural based on the Virtual Ambisonics Approach. In this case, it is also possible to perform an environmental simulation in the Ambisonics domain (other software is needed for this), and then to convert it to binaural. This is definitely the best simulation heard for sources located between one metre and 15 centimetres. Nevertheless, the lack of envi-	//

6 *See* www.ircam.fr/equipes/salles/listen/

				ronmental simulation makes it unsuitable for many tasks such as the simulation of sound sources placed in real environments.	

Appendix E

Notes on the CD submitted with the Thesis

Attached to this dissertation is a CD containing different files; it represents an "excerpt" of the work performed during this Ph.D. research.
Here follows a list of the files found in the different folders within the CD itself:

1_Binaural_Recordings

This folder contains four binaural recordings carried out with the dummy head used in this research project.

2_Pete_Batchelor

This folder contains two stereo versions (one stereo down-mix and one converted to binaural from twelve channels; *see* Section 9.2 of the thesis) of the piece *Kaleidoscope: Arcade* by Pete Batchelor. The two versions of the piece are published on the CD *Reflections* by Pete Batchelor (2008).

3_Offline_Applications

This folder contains a beta version (with limited functionalities, merely for demo purposes) of the offline Source Positioning Binaural application (*see* Section 7.3.1).
In this folder may also be found an implementation of a "moving sound sources" binaural application. As pointed out in the Introduction to this thesis and in Chapter 7, at the beginning of this research work the author, together with the supervision team, decided not to implement any HRIR interpolation technique; instead, to perform the simulation of moving sound sources in the Ambisonic domain. Nevertheless, a possible implementation of moving sound sources simulation directly in the binaural domain has been attempted using a simple sinusoidal wave for interpolating between the different HRIRs within the trajectory of the source. It is essential to underline that this is simply a beta version, a non-tested and non-debugged application that should serve only as an example of a possible implementation of such techniques.
For information about how to launch these applications, *see* Appendix F.

4_C++_Code

This folder contains the two C++ code files (*000_Convolve_0.92.cpp* and *000_ConvolveMov_0.5.ccp*) corresponding to the two offline applications (*see* previous point). In order to compile these files, it is essential to install the *soundfile++*[1] library and the *fft2*[2] C++ function, plus, obviously, a C++ compiler (such as *gcc* for Unix machines).

5_AmbiTOBin_RealTime

This folder contains a beta version (with limited functionalities, merely for demonstration purposes) of the real-time MaxMSP Ambisonic-to-binaural application (*see* Section 7.4). It also contains the electronic version of a short manual of the application itself (*see* Appendix G). In order to run the application, the folder needs to be copied onto the hard drive of the computer to be used (the software will work only on Intel Mac machines), in which MaxMSP 4.6 needs to be installed; either the full or runtime version should be adequate. Then, it should be sufficient to launch the file named *000_AmbiTOBin_RealTime* in order to run the application.

6_Other_Applications

This folder contains three MaxMSP applications that have been developed during the Ph.D. research period. The three applications have been organized into three separate folders: *Distance_Simulator* (*see* Section 6.3.4), *Test_Platform_I*, and *Test_Platform_II* (the two platforms developed for the two perceptual tests: *see* Chapter 8). In order to run the applications, each of the folders needs to be copied onto the hard drive of the computer to be used (the software will work only on Intel Mac machines), in which MaxMSP 4.6 needs to be installed; either the full or runtime version should be adequate. Then, it should be sufficient to launch one of the three files with a name starting with *000_* in order to run the application.

1 *See* http://soundfile.sapp.org/

2 Written by Dale Carstensen, the Antares project, Los Alamos National Laboratory, 16 March 1981 for Unix version 6.

Appendix F

Short manual for the offline applications
Versions Beta 0.92 and Beta 0.5

For the design overview of this piece of software, *see* Chapter 7 of the thesis.

In order to run it, the folder needs to be copied onto the hard drive of the computer to be used (the software will work only on Intel Mac machines, and for a correct functioning of both pieces of software it is recommended that the computer has a minimum of 2Gb of RAM). Using the Terminal, and entering the folder that has been copied, two files can be opened: *000_Convolve_0.92*, which will launch the Source Positioning Binaural application, and *000_ConvolveMov_0.5*, which will open the "moving sound source" binaural application.

For both applications, a list of instructions will appear in the Terminal window: the instructions need to be followed carefully in order for the application properly to spatialize the required signals. Spatializing longer audiofiles will, obviously, require a longer time (up to several hours for the reverb simulation).

As an example, in the following lines is reported the series of instructions and parameters to be input by the user (these are reported in bold) when using the offline Source Positioning Binaural application for spatializing a mono audiofile with the name *Audiofile.wav* at 90° of azimuth, 0° of elevation, with low frequencies compensation calibrated at 100 Hz and with the following coefficients for the weighted sum: 0dB (direct signal), -6 dB (early reflections), -12 dB (reverb) and -6 dB (low frequencies compensation).

- *OPEN THE TERMINAL*
- ***cd .../3_Offline_Applications***
- ***.../000_Convolve_0.92***
- *Please, insert the path and the name of the audiofile that needs to be spatialized. The audiofile needs to be a .wav file (Microsoft Wave), 44100 Hz and 16 bits. If the audiofile is in the same folder of the convolve software, just put the name of it:*
- ***.../Audiofile.wav***

- *Now, please, insert the degrees of azimuth (measured anti-clockwise and rounded to 2.5 degrees, please use only positive values) and elevation (rounded to 10 degrees) where you want your file to be spatialized, separated by a blank space:*
- ***90 0***
- *Please decide the type of environment you want to be simulated: 1 for a large environment, 2 for a medium environment and 3 for a small environment:*
- ***2***
- *Finally, decide the cut-off frequency for the lowpass filter used for the low frequencies compensation algorithm. The value needs to be rounded at 10Hz and between 80Hz and 250Hz:*
- ***100***
- *The state of the processing for the direct components is: 100%*
- *The state of the processing for the early reflections components is: 100%*
- *The state of the processing for the reverberant components is: 100%*
- *The state of the processing for the low frequencies compensation is: 100%*
- *THE CONVOLUTION HAS BEEN EXECUTED CORRECTLY! THE SPATIALIZED AUDIOFILES HAVE BEEN SAVED IN THE SAME FOLDER OF THE PROGRAM WITH THE FOLLOWING NAMES:*
  - *"outputDirect.wav" for the direct signal stereo audiofile*
  - *"outputEarlyRef.wav" for the early reflections signal stereo audiofile*
  - *"outputRev.wav" for the reverberant signal stereo audiofile*
  - *"outputLC.wav" for the low frequencies compensation signal stereo audiofile*
- *It is possible now to have a fifth stereo audiofile, which will contain a weighted mix of the first four (of course, it is possible to choose the weight for all the four signals).*

  Do you want to proceed with the processing (Y/N)?
- ***Y***

- *Ok, now give the value of the weight in dB (0 dB is the maximum, and corresponds to the signal with no reduction) for the direct signal:*
- ***0***
- *And now the weight in dB for the early reflections signal:*
- ***-6***
- *And now the weight in dB for the reverberant signal:*
- ***-12***
- *And finally the weight in dB for the low frequencies compensation signal:*
- ***-6***
- *The state of the processing for the weighted sum is: 100%*
- *The weighted sum has been executed correctly, and the signal has been saved in the audiofile "outputSum.wav" in the same folder of the program!*
- *THE PROGRAM IS NOW FINISHED!*

Finally, in the following lines is reported the series of instructions and parameters to be input by the user (again, these are reported in bold) when using the offline "moving sound source" binaural application for spatializing a mono audiofile with the name *Audiofile.wav* between 90° and 180° of azimuth in 5 seconds.

- *OPEN THE TERMINAL*
- ***cd .../3_Offline_Applications***
- ***.../000_ConvolveMov_0.5***
- *Please, insert the path and the name of the audiofile that needs to be spatialized. The audiofile needs to be a .wav file (Microsoft Wave), 44100 Hz and 16 bits. If the audiofile is in the same folder of the convolve software, just put the name of it:*
- ***.../Audiofile.wav***
- *Now, please, insert the degrees of azimuth (measured anti-clockwise and rounded to 2.5 degrees, please use only positive values) for the starting and for the end points where you want your file to be spatialized, separated by a blank space:*
- ***90 180***

- *Finally, insert the duration in seconds of the movement you want to simulate from the two points given before:*
- ***5***
- *THE CONVOLUTION HAS BEEN EXECUTED CORRECTLY! THE SPATIALIZED AUDIOFILE HAS BEEN SAVED IN THE SAME FOLDER OF THE PROGRAM WITH THE NAME "outputMov.wav"*
- *THE PROGRAM IS NOW FINISHED!*

Appendix G

Short manual for the real-time application
Version Beta 0.2

For the design overview of this piece of software, *see* Chapter 7 of the thesis.
For information about how to launch the application, *see* Appendix E.
In order to run it, the folder needs to be copied onto the hard drive of the computer to be used (the software will work only on Intel Mac machines), in which MaxMSP 4.6 needs first to be installed; either the full or runtime version should be adequate. Then, it should be sufficient to launch the *000_AmbiTOBin_RealTime* file in order to run the application.
Here follows a list of the different sections and their functions of the MaxMSP application:

- **Source** section (blue). In this section the parameters for the sound source may be set. In the menu under *Select File* the sound source signal between an anechoic guitar, an anechoic speech or a synthesized *ping* noise may be selected. Pressing the button at the right of *or open mono file*, any monophonic audiofile to be spatialized may be imported. Below that, it is possible to choose the azimuth and elevation of the spatialized sound source. Finally, on the left part of the section, there is a toggle for the start/stop operation, and another for looping the playback.
- **Audio ON/OFF** section (red). The main control of this section is the audio ON/OFF button (on the left of the *AUDIO ON/OFF* writing): when this button is activated, the whole audio system will be turned on (it therefore needs to be activated before the application is begun). Within this section it is also possible to choose the orientation of the head in degrees of Azimuth (the relative position of the sound source will then change), and to decide whether to listen either to the binaural spatialized signal or to the stereo signal by activating one of the two buttons; this option is provided in order to be able to draw comparisons between binaural and standard stereo spatialization modes.
- **Low frequency compensation** section (green). In this section it is possible to choose the cutoff frequency for the low-pass filter and also the level of the filtered non-spatialized signal to be summed to the spatialized signal, in order to recover the

low frequencies lost in the binaural spatialization process. Through placing the cursor on the equalization diagram, the gain, frequency, and Q-factor for the equalization filter itself may be changed.

- **Ambisonic encoding and decoding** section (green). The only function of this section is to monitor the levels of the signals in the different stages of the spatialization, therefore the levels of the nine 2nd Order Ambisonic channels, of the twelve channels of the dodecahedron loudspeaker's setup for the direct signal spatialization, and the four channels of the square loudspeaker's setup for spatialization of the early reflections and reverberant signals.
- **IMPORTANT** section (blue). This section contains a highly important announcement about how to set up the vector sizes of both the I/O and the signal in order correctly to run the application. Furthermore, this section allows the changing of these parameters and the monitoring of the CPU Utilization percentage (without having to enter the *DSP status* menu). The recommended settings in order to run the application without problems are:
  - I/O Vector Size: 512 samples
  - Signal Vector Size: 256 samples
- **Master Faders** section (red). Here, it is possible to change the levels for the direct, early reflections and reverberant signal components, together with the master output level. As pointed out under the section title, when moving the left fader, the right one will move, too; in contrast, while moving only the right fader, the left fader will be left unchanged. On the sides of all of the faders the levels of the signals may be monitored. For master-level monitoring, the monitoring mode may be selected from a menu. For the early reflections and the reverberant components, moving the horizontal fader will change the pre-delay in ms. Below the individual component faders, a button is present for resetting the levels and the pre-delay times for direct, early reflections, and reverberant signals (without this resetting the master-level faders).